

DOCUMENT RESUME

ED 435 632

TM 030 231

AUTHOR Shinkfield, Anthony J.; Stufflebeam, Daniel L.
TITLE Teacher Evaluation: Guide to Effective Practice. Evaluation in Education and Human Services.
INSTITUTION Center for Research in Educational Accountability and Teacher Evaluation (CREATE), Kalamazoo, MI.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO ISBN-0-7923-9674-x
PUB DATE 1995-00-00
NOTE 408p.; "With contributions by: Carol Anne Dwyer, Sandra Horn, Madeline Hunter, Edward Iwanicki, Richard Manatt, Thomas McGreal, Robert Mendro, William Sanders, William Webster, Graeme Withers."
CONTRACT R117Q00047
AVAILABLE FROM Kluwer Academic Publishers, 101 Philip Drive, Assinippi Park, Norwell, MA 02061.
PUB TYPE Books (010) -- Guides - Non-Classroom (055)
EDRS PRICE MF01/PC17 Plus Postage.
DESCRIPTORS *Administrators; *Criteria; *Educational Administration; Elementary Secondary Education; *Models; *Teacher Evaluation

ABSTRACT

This guide to effective teacher evaluation is organized around the core issues of professional standards, a guide for applying the Joint Committee's "Standards," 10 alternative models for the evaluation of teacher performance, and an analysis of these 10 models. The chapters are: (1) "Historical Perspectives of Teacher Evaluation"; (2) "Standards and Criteria for Teacher Evaluation"; (3) "School Professionals' Guide to Improving Teacher Evaluation Systems"; (4) "Models for Teacher Evaluation"; and (5) "An Analysis of Alternate Models." Each chapter contains references. (Contains 11 figures and 11 tables.) (SLD)

TEACHER EVALUATION:

Guide to Effective Practice

Anthony J. Shinkfield

Daniel Stufflebeam

ED 435 632

TM030231

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- ☒ This document has been reproduced as received from the person or organization originating it.
- ☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Kluwer Academic Publishers

Boston / Dordrecht / London

TEACHER EVALUATION:

Guide to Effective Practice

Evaluation in Education and Human Services

Editors:

George F. Madaus, Boston College,
Chestnut Hill, Massachusetts, U.S.A.
Daniel L. Stufflebeam, Western Michigan
University, Kalamazoo, Michigan, U.S.A.

Other books in the series:

Madaus, G. and Stufflebeam, D.:
Education Evaluation: Classic Works of Ralph W. Tyler
Gifford, B.:
Test Policy and Test Performance
Osterlind, S.:
Constructing Test Items
Smith, M.:
Evaluability Assessment
Ayers, J. and Berney, M.:
A Practical Guide to Teacher Education Evaluation
Hambleton, R. and Zaal, J.:
Advances in Educational and Psychological Testing
Gifford, B. and O'Connor, M.:
Changing Assessments
Gifford, B.:
Policy Perspectives on Educational Testing
Basarab, D. and Root, D.:
The Training Evaluation Process
Haney, W.M., Madaus, G.F. and Lyons, R.:
The Fractured Marketplace for Standardized Testing
Wing, L.C. and Gifford, B.:
Policy Issues in Employment Testing
Gable, R.E.:
Instrument Development in the Affective Domain (2nd Edition)
Kremer-Hayon, L.:
Teacher Self-Evaluation
Payne, David A.:
Designing Educational Project and Program Evaluations
Oakland T. and Hambleton, R.:
International Perspectives on Academic Assessment
Nettles, M.T. and Nettles, A.L.:
Equity and Excellence in Educational Testing and Assessment

TEACHER EVALUATION: Guide to Effective Practice

by
Anthony J. Shinkfield
and
Daniel L. Stufflebeam

with contributions by:

Carol Anne Dwyer
Sandra Horn
Madeline Hunter
Edward Iwanicki
Richard Manatt
Thomas McGreal
Robert Mendro
William Sanders
William Webster
Graeme Withers



KLUWER ACADEMIC PUBLISHERS
Boston/Dordrecht/London

Distributors for North America:
Kluwer Academic Publishers
101 Philip Drive
Assinippi Park
Norwell, Massachusetts 02061 USA

Distributors for all other countries:
Kluwer Academic Publishers Group
Distribution Centre
Post Office Box 322
3300 AH Dordrecht, THE NETHERLANDS

Library of Congress Cataloging-in-Publication Data

Shinkfield, Anthony J.

Teacher evaluation : guide to effective practice / by Anthony J.
Shinkfield and Daniel L. Stufflebeam : with contributions by Carol
Anne Dwyer ... [et al.].

p. cm. -- (Evaluation in education and human services)

Includes bibliographical references and indexes.

ISBN 0-7923-9674-X (acid-free paper)

1. Teachers--Rating of. I. Stufflebeam, Daniel L. II. Title.

III. Series.

LB2838.S53 1995

371.1'44--dc20

95-14669

CIP

Copyright © 1995 by Kluwer Academic Publishers. Second Printing 1997.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, mechanical, photo-copying, recording, or otherwise, without the prior written permission of the publisher, Kluwer Academic Publishers, 101 Philip Drive, Assinippi Park, Norwell, Massachusetts 02061.

Printed on acid-free paper.

Printed in the United States of America

Contents

ACKNOWLEDGMENTS	ix
ABOUT THIS BOOK	1
Foreword	1
An Overview of Contents	2
Audiences	6
Contributors	8
1	
HISTORICAL PERSPECTIVES OF TEACHER EVALUATION	9
Pre-World War II	11
Post-World War II Until The Mid-1970s	13
The Late 1970s To The Present	23
Future Challenges In Evaluation	30
A National Center for Research on Teacher Evaluation and Dissemination of Outcomes	34
The Praxis Series: Professional Assessments for Beginning Teachers	35
The National Board for Professional Teaching Standards	36
Conclusion	37
References	38
2	
STANDARDS AND CRITERIA FOR TEACHER EVALUATION	43
Preamble: The Place and Importance of Standards	43
References	47

Professional Standards for Assessing and Improving Teacher Evaluation Systems	47
Criteria for Performance-Based Teacher Assessments:	
Validity, Standards, and Issues, <i>by Carol Anne Dwyer</i>	62

3
SCHOOL PROFESSIONALS' GUIDE TO IMPROVING TEACHER EVALUATION SYSTEMS

	81
Preamble: A GUIDE to Improving Teacher Evaluation Systems	81
Introduction	82
Teacher Evaluation: Its Purpose, Meaning, and Improvement	85
Organizing a Participatory Project to Improve Teacher Evaluation	102
Profiling the Current Teacher Evaluation System	106
Determining which Personnel Evaluation Standards Are Met by the Present Evaluation System	111
Deciding and Planning How to Improve the Teacher Evaluation System	117
Improving the Current Teacher Evaluation System	121
Planning and Implementing the Evaluation Improvement Project	127
References	131
Form for Documenting a Teacher Evaluation System	133
Questions to Be Answered in Addressing the Personnel Evaluation Standards	153

4
MODELS FOR TEACHER EVALUATION

	173
Preamble: Overview of Alternative Models	173
Madeline Hunter: Instructional Effectiveness Through Clinical Supervision	187
Thomas McGreal: Characteristics of Successful Teacher Evaluation	208
Appendix: An Example of an Evaluation System That Reflects the Commonalities of Successful Systems	230
Edward Iwanicki: Contract Plans—A Professional Growth-Oriented Approach to Evaluating Teacher Performance	245
Getting Value from Teacher Self-Evaluation, <i>by Graeme Withers</i>	261
Richard Manatt: Teacher Performance Evaluation	271
Toledo School District: Intern and Intervention Programs	289
Principal and Peer Evaluation of Teachers for Professional Development, <i>by Anthony Shinkfield</i>	302
The National Board for Professional Teaching Standards: Assessing Accomplished Teaching	319

CONTENTS

vii

The Tennessee Value-Added Assessment System (TVAAS): Mixed Model Methodology in Educational Assessment, by <i>William L. Sanders and Sandra P. Horn</i>	337
An Accountability System Featuring Both "Value-Added" and Product Measures of Schooling, by <i>William Webster and Robert Mendro</i>	350
5	
AN ANALYSIS OF ALTERNATE MODELS	377
Introduction	377
A Summary of the Purposes of the Ten Models Selected in Chapter 4	378
An Examination of the Models Against the Joint Committee's <i>Standards</i>	385
The Worth of the Presented Models for Decision Making	392
References	394

Acknowledgments

This work was supported by the U.S. Department of Education, Office of Educational Research and Improvement (OERI) through grant No. R117Q00047 to Western Michigan University for the work of the National Center entitled Center for Educational Accountability and Teacher Evaluation (CREATE). The opinions expressed are those of the authors, and do not necessarily reflect those of the U.S. Department of Education.

ABOUT THIS BOOK

Foreword

Formal teacher evaluation has long been considered important by the public. However, until recently most schools evaluated teachers in only the most cursory manner, e.g., by the principal's annual brief observation of the teacher's classroom performance. The last 15 years has seen a dramatic development of the technology of teacher evaluation. Also, there has been much more attention in the schools to making teacher evaluation extensive, systematic, and valid. Much of this growth arose from state mandates and from educational institutions whose leaders seek improved means of evaluating teachers. Closely connected to these worthwhile enterprises has been the often expressed need to elevate the respect accorded to teacher evaluation by using professional standards. This book is an attempt to present and examine important developments in teacher evaluation, show their interrelationships, and offer practical guidelines for using teacher evaluation models.

We organized the book around 4 dominant, interrelated core issues: professional standards, a GUIDE for applying the Joint Committee's *Standards* (which are featured in Chapter 3), 10 alternative models for the evaluation of teacher performance, and an analysis of these selected models. The book draws heavily upon the research and development conducted by the federally funded national Center for Research on Educational Accountability and Teacher Evaluation (CREATE).

We hope that the reader will grasp the essence of the experience of sound teacher evaluation and apply its principles, facts, ideas, processes, and procedures. To this end, we offer information that is as up-to-date, and practically useful, as possible. For example, Chapter 3 presents a thoroughgoing, logical sequence of steps to guide readers in examining teacher evaluation systems for adherence to definitive pro-

fessional standards. Moreover, the book invites and assists school professionals and other readers to examine the latest developments in teacher evaluation.

Evaluation, both formal and informal, is inextricably interwoven with the entire process of education. Since the quality of learning depends largely on the quality of teaching, teacher evaluation clearly is essential in effective schools. Administrators, other leaders in education, and school communities have long realized the often disturbing truth of this situation. Some schools and districts have failed to plan and carry out an appropriate teacher evaluation scheme because they could not locate a suitable model. Others failed because they lacked the knowledge or will to make a sound model work in the face of difficulties. By addressing these kinds of issues, this book should guide readers toward attaining important goals in teacher evaluation. These include improving the performance of teachers, students, and the organization as a whole.

An Overview of Contents

The book has five chapters. Together they explain and amplify the four main cores already listed.

Chapter 1 offers a historical perspective of teacher evaluation. Our literature search brought us to the realization that extensive material on this topic does not exist (and certainly not in any consolidated form). Formal teacher evaluation is largely a recent phenomenon. We hope, therefore, that this section is a useful, interesting contribution to the emerging field of assessment of teacher performance.

Chapter 1 first examines pre-World War II ideas of teacher evaluation. The common view was that students were responsible for their learning. Accordingly, schools attributed learning deficits to the student rather than the teacher.

However, during the Victorian era, England initiated the first coordinated, nationwide program to assess teachers and reward them accordingly. This was called payment by results. This led to the English inspectorial system that has persisted in that country, and others. Many annual reports and other sources reveal that, as one would expect, principals were conducting informal assessments of teachers in the U.S. during the 1900s.

From the conclusion of World War II until the mid-1970s, the literature and research reports reveal a growing consensus in the major purposes of teacher evaluation. The predominant factor, it seems, was the growing belief that the entire educational system must gain from improved teacher performance arising from widely acceptable evaluation processes. During this time states and school districts made tentative beginnings in areas like systematic accountability of teachers and appraising teachers based on the learning of their students.

The final section of Chapter 1 treats the late 1970s to the present. It looks into the emerging ideas of evaluation and accountability, teacher certification, and the legal and political aspects of evaluation.

Chapter 1 concludes with a view of future challenges in evaluation. Formal teacher evaluation may be in its infancy, but the growth rate has been most marked, particularly during the past decade. Recent historical perspectives necessarily reference contrasting models for evaluating teachers; the importance of using professional standards to assess and improve teacher evaluation systems; CRE-ATE's extensive, ongoing, nationwide studies of teacher evaluation; and the emergence of an annual national teacher evaluation institute. Thus, in the 1980s and 1990s there has been a considerable advance in the theory and practice of this important aspect of education.

Chapter 2 centers on standards and criteria for teacher evaluation. If the evaluation field is to achieve its potential contribution in any area, it must yield dependable assessments of all aspects of a discipline or system. It follows that evaluators must develop and apply professional standards to help ensure that all aspects of evaluation attain the highest levels of fairness and quality.

Chapter 2 also discusses **Professional Standards for Assessing and Improving Teacher Evaluation Systems**. It outlines the major undertaking by a widely-representative national Joint Committee during the 1980s culminating in the publication of *The Personnel Evaluation Standards* in 1988. These *Standards* provided education with a widely endorsed set of guiding principles. The rigorous application of the *Standards* strengthens and adds credibility to systems and practices of personnel evaluation and protects teachers and others from corrupt evaluation practices. They should also help to mitigate evaluation-related conflicts among different interest groups. The chapter discusses development of the Joint Committee's *Standards*. It delineates areas where they are applied. Then it provides a brief introductory statement about each of the four basic principles of sound evaluation—propriety, utility, feasibility, and accuracy.

Professional standards are the foundation for this book. They play a remarkably pervasive role, directly and indirectly in every chapter. We believe that school professionals take a serious view of teacher evaluation only if they are knowledgeable about the Joint Committee's *Standards*. Chapter 3, which is a user's guide to assessing and improving teacher evaluation systems, and the Preamble to Chapter 4, an overview of alternative models, exemplify this point most markedly.

In **Chapter 2** Carol Anne Dwyer writes from her recent experiences in leading the development of Praxis (the Educational Testing Service successor to NTE). Her concern is the historic lack of standards for assessing teacher competence. Thus, her article's emphasis on criteria for teacher evaluation relates strongly to the emphasis placed on professional standards for evaluations in Chapter 2. She observes that PRAXIS, inter alia, keys to 3 different aspects of the pedagogy:

content-specific pedagogical knowledge, knowledge of general principles of teaching and learning that transcends different disciplines, and the application of this knowledge and skills to actual classrooms. These require different assessment methods. Dwyer also summarizes the recent progress in improving performance assessment. She documents progress in closing the gap between standards for sound performance assessment and actual practice.

This article is on the leading edge in both defining and addressing issues in the validation of teacher assessments. It also emphasizes the complexity and considerable expense involved in defining and applying assessment standards.

Chapter 3 is the **School Professionals' GUIDE to Improving Teacher Performance Evaluation Systems**. In many ways, this chapter is a companion document to Chapter 2, particularly in respect to the Joint Committee's *Standards*. This part shows how the *Standards* can be effectively and systematically used to examine extant or contemplated teacher evaluation systems.

Chapter 3 begins with a discussion of which particular standards are most important in assessing the adequacy of a teacher evaluation system. It next presents a conceptual framework delineating factors that define and influence such systems. Advice is offered on how to use this framework to examine systems. The chapter concludes with recommendations on how to organize a participatory project involving all stakeholder groups.

These activities in the **GUIDE** lead to formulating a new system for teacher evaluation or improving an existing one. The **GUIDE** refers to the *Standards* in discussing these processes. The **GUIDE** concludes with 3 appendices. These offer highly relevant exercises plus a table of contents for a manual on teacher evaluation.

We hope that the **GUIDE** will prove very beneficial to school districts. Its authors designed it to be easy to understand, easy to use, and to apply to evaluation needs in schools. It recommends and provides direction for involving all stakeholders in improving a district's teacher evaluation system.

Chapter 4, Models for Teacher Evaluation, has two main objectives. The first is to give a general overview of teacher evaluation models. The second is to present 10 models that are widely used or especially interesting.

In the preamble to Chapter 4, we explain our intended meaning of the term model. In effect, we redirect emphasis from various ideas of teacher evaluation that portray the actual process of evaluation to those that prescribe a preferred, ordered set of steps for conducting teacher evaluations. The presented models vary from being highly directive to less directive. However, they address a common purpose. It is to evaluate teachers so well that there are clear and practical benefits for schools, school districts, and students.

We based our overview of teacher evaluation models on a listing provided by Michael Scriven in his **CREATE** publication called **TEMP Memo 2** (September 1991). We added self-evaluation to the list and briefly outlined each model. The 15

models contained in the overview are Traditional Impressionistic, Clinical Supervision, Research-Based Checklist, High Inference Judgments, Interviewing, Paper and Pencil Tests, Management by Objectives, Job Analysis, Duties-Based Approach, Theory-Based Approach, Student (Learning) Improvement Outcomes, Consumer Ratings, Peer Ratings, Self-Evaluation, and Metaevaluation of Existing Models.

The 10 selected models that comprise the remainder of Chapter 4 cover most of the listed TEMP Memo approaches. We believe that they show the range of models available for school professionals to consider.

We define *formative* evaluation as systematically assessing the merit and worth of some enterprise to guide its continuing revision during the process. We define *summative* evaluation as a comprehensive assessment of the merit and worth of the enterprise, including especially its outcomes at the end of the process. The remainder of Chapter 4 depicts 10 evaluation models divided into (1) formative, (2) formative and summative, and (3) summative approaches to teacher evaluation.

The first four models, Chapters 4.1–4.4, are essentially formative: Madeline Hunter's *Instructional Effectiveness Through Clinical Supervision*, Thomas McGreal's *Characteristics of Successful Teacher Evaluation*, Edward Iwanicki's *Contract Plans - A Professional Growth-Oriented Approach to Evaluating Teacher Performance*, and Graeme Withers' *Getting Value from Teacher Self-Evaluation*.

The next three models, Chapters 4.5–4.7, are both formative and summative. They include Richard Manatt's *Teacher Performance Evaluation*, Toledo School District's *Intern and Intervention Programs*, and Anthony Shinkfield's *Principal and Peer Evaluation of Teachers for Professional Development*.

The final three models, Chapters 4.8–4.10, are summative. They include the National Board for Professional Teaching Standards' work on *Assessing Accomplished Teaching*, William Sanders and Sandra Horn's *The Tennessee Value-Added Assessment System (TVAAS)—Mixed Model Methodology in Educational Assessment*, and William Webster and Robert Mendro's *An Accountability System for School Improvement*.

Chapter 5 analyzes the 10 models presented in Chapter 4. It tries to achieve the following:

- A summary display of the purposes of each model, which is placed under one of three headings: formative, formative and summative, and summative.
- An examination of the models against the Joint Committee's *Standards* to find their main strengths and weaknesses
- Our discretionary value judgment about the worth of these models considered against the main uses of teacher evaluation models for decision making.

Taken together, these elements comprise a useful consumer's guide. It stems logically from the development of the first three of this book's four main cores: standards for teacher evaluation, the GUIDE to improving teacher evaluation

systems by applying the Joint Committee's *Standards*, and the presentation of alternative models for teacher evaluation.

Audiences

In the improvement of teacher evaluation, it is critical to organize collaborative work by those involved in the evaluation process and those influenced by its outcomes. Almost without exception the approaches to teacher evaluation given in this book bear this out. Most of the approaches either state or imply the need for collaboration in their models and guidelines for their use.

Thus, we wrote the book purposely to provide useful guidance to professional personnel connected with schools—particularly teachers, principals and other school administrators, and superintendents and board members. We recommend that officials in state education departments and professional education organizations use this book to develop better state and district systems of teacher evaluation. Parents, and the public generally, are also stakeholders in the products of the educational process. Many of them should find the book useful for assessing their school districts' teacher evaluation practices. Concerted actions by all the stakeholders to improve teacher evaluation should lead to improved public credibility of the schools and the education establishment.

Teacher evaluation is growing in importance both quickly and pervasively. We therefore recommend the book to those whose working lives could see them involved with teaching, or evaluation, or both. These include education professors and college and university students at both undergraduate and graduate levels.

There is a strong reason for research to continue in teacher evaluation because it is an emerging area of national interest and need. Any group of professionals using the GUIDE presented in Chapter 3, where we assess models in the light of accepted standards, will discover the imperfections of any model. The outcomes of sound, focused, research studies should help improve many aspects of teacher evaluation practice. These include validity, potential utility, feasibility, consonance with human rights legislation, and the effective and efficient application of validated models.

Many people, therefore, have a legitimate interest in teacher evaluation. These include:

- Those who are immediately involved in teacher evaluation—to lists already given, students are added because the basic purpose of any program within education is to improve student learning.
- Those who provide leadership for improving teacher evaluation practices at the state and national levels

- Parents and the wider public whose educational accountability requirements should be met
- Those persons responsible for teacher training and student teachers themselves
- Evaluation researchers

Considered as a whole, this is a widely disparate group, often holding different views about education and its many complex functions. Therefore, we wrote the book straightforwardly in the hope that it will be readable, interesting, and practicable.

Responsibilities and Reassurances

A discussion of teacher evaluation often leads to the question of what other persons should be evaluated. That is, who else should communities hold accountable for producing quality outcomes from educational processes? There is a growing recognition that the acceptance of a teacher evaluation system is strengthened if all administrators in the district (superintendent, principals, vice-principals, and perhaps school board members could be added) are also evaluated, and particularly so if common principles and procedures are used. Someone must evaluate the evaluators. Otherwise, the public and other stakeholders can have no assurance that the evaluators are employing sufficient accountability measures to safeguard student interests and motivate educational improvements. There is also growing realization that school districts must evaluate support and special staff and that the process should not differ significantly from that for teachers and administrators. This country is a considerable distance from this happening universally, and even further from it occurring in highly credible and acceptable ways.

We did not address these important adjunct areas for personnel evaluation in this book. However, CREATE is developing methodology designed to give very considerable direction to evaluation of administrators, support personnel, and board members.

The major stakeholders should share responsibility for developing teacher evaluation in a school or district. The major actors include the teachers, administrators, board members, and teachers' organization. However, responsibility for carrying out the process satisfactorily rests with different groups, or individuals, according to the decision-making involved. For instance, administrators will clearly strengthen a district and its schools by doing a thoroughly professional job of assessing and selecting newcomers. Also, principals are largely responsible for using best possible formative evaluation for staff professional development. Unfortunately, evaluations associated with tenure, promotion, and reassignment are often haphazard and often lack the credibility needed to guide and defend personnel

decisions. The board must enact policies to assure that the district professionally evaluates teachers and other district personnel to benefit the district and its students.

As professional people, teachers themselves must engage in evaluation for both professional development and accountability. Regular external evaluation acts as a supplement to self-evaluation, an ongoing process essential for any professional. Teacher involvement in developing evaluation models is important. The school teacher, school administrator, and school board member can assure themselves and their clients that accountability is being practiced properly only when they conduct, report, and act on the results of systematic, valid evaluation.

Contributors

We owe a great debt to those who have made this book possible. We refer to those who have allowed us to present a version of their model for teacher evaluation: the late Madeline Hunter, Richard Manatt, Thomas McGreal, and Edward Iwanicki. We hope that we have done justice to the Toledo School District and the National Board for Professional Teaching Standards in our discussion of their models. Thanks to Carol Anne Dwyer, Graeme Withers, William Sanders and Sandra Horn, and William Webster and Robert Mendro for allowing us to reproduce their articles; to Thomas McGreal for permission to print in full the appendix to his book, *Successful Teacher Evaluation*; and to Daniel Stufflebeam's coauthors David Nevo, Bernard McKenna, and Rebecca Thomas for agreeing to our inclusion of a version of the *School Professionals' GUIDE*.

We especially appreciate the support given to us by the U.S. Department of Education, Office of Educational Research and Improvement (OERI) through a grant to Western Michigan University for the work of CREATE. The opinions expressed in this book do not necessarily reflect those of the U.S. Department of Education. This book helps to fulfill one of CREATE's chief missions, which is to extend knowledge of the place and importance of teacher evaluation and to offer guidance about the use of models. Finally, we very much appreciate the ready and most able assistance given us by the book's reviewers. CREATE secures external evaluation of all CREATE products by appropriately qualified people. Graeme Withers, Senior Research Officer with the Australian Council for Educational Research, and Bernard McKenna, retired official of the National Education Association, fulfilled this task in a thoroughly professional manner. We appreciate their significant contributions.

— Anthony J. Shinkfield

— Daniel L. Stufflebeam

1

HISTORICAL PERSPECTIVES OF TEACHER EVALUATION

Over the ages, teachers have always been evaluated. Socrates' pupils undoubtedly had opinions about his teaching skills in the 5th century B.C. Tom Brown, of *Tom Brown's School Days*, certainly made clear his impressions of the effectiveness of his mid-Victorian English grammar school teachers. Most parents today know what their children think of their teachers. The fact that any of these opinions may be far from the truth does not, and will not, prevent their expression. The trouble with teacher evaluation is that teaching itself is a highly complicated process. No one knows precisely what ideal role a teacher should perform to affect excellent student learning, not even when the context of a classroom is specified.

Whether it was the inherent difficulties of teacher assessment or the assumption that teachers were infallible whereas students were responsible for their own learning, formal evaluation of teachers was virtually unknown until the turn of the 20th century. Even thereafter, for the next half century or more, very few schools and school districts attempted formal processes to gauge the work of teachers.

Movements commenced in the 1970s, and considerably increased in the '90s, have given an abrupt, and significant, impetus to teacher evaluation models and approaches. One such catalyst has been the 1983 federal government's report, *A Nation at Risk: The Imperative of Education Reform*. Although uncertainties about many aspects of teacher evaluation prevail, its importance is acknowledged by the adoption of relevant policy and practice documents by almost all school districts. Such decisions are most often motivated as much by the enactment of state legislative requirements as the desire to improve the professional status of teachers.

There are indications, in many states, that evaluations have become regulatory, linked more often than not to strengthening the state's control of teachers.

Collective bargaining agreements have further politicized the nature of teacher evaluations. Contracts specify policies, procedures, reporting and, when negative results arise, remediation in line with due process rules. Sometimes contracts specify evaluative criteria and instrumentation. All too often, this creates gridlock. Thus, both the union and the school district are constrained to use a standardized approach that does not and cannot reflect the diversity of student needs and teacher assignments seen across the span of classrooms. Such processes, with emphasis being given to judgments based on organizational requirements conforming to agreed-upon rules, are summative. These evaluations are influenced by the political climate of the various elements of a school district and numerous other variables including the size and structure of the system.

Nonetheless, there is ample evidence to suggest that a formative influence is present in many of the approaches seen in operation today. Despite the difficulty of defining teaching and its subtleties, diversities, and effectiveness for student learning, most teacher evaluation schemes in use today at least purport to have teacher competency enhancement as one of their main objectives.

Anyone reviewing teacher evaluation at the turn of the next century will comment on the very considerable advance in the theory and practice of this important aspect of education during the 1980s. It is hoped that by then many of the present problems of teacher evaluation will have been resolved. It seems likely that the publication of *The Personnel Evaluation Standards* by the Joint Committee on Standards for Educational Evaluation in 1988 will have played a significant part in the advancement of the purposes and practices of teacher evaluation.

It is difficult, if not impossible to place the historical perspectives of teacher evaluation into neat boxes. It was not until the 1950s that any serious writing was undertaken in personnel evaluation, and that pertaining to teachers lagged still a further decade behind. Moreover, although it was not difficult to find various instances of teacher evaluation being practiced in the western world, a coherent body of theory did not begin to emerge until the 1960s.

In a quite arbitrary fashion, we have divided this chapter into four sections: Pre-World War II, Post-World War II until the mid-1970s, the late 1970s to the present, and the emergence of standards to give both a framework and legitimacy for approaches to the evaluation of teachers.

Pre-World War II

During the period from the late 18th century to the mid-20th century, one would need to leap from continent to continent to find any instances of what might vaguely be termed attempts to evaluate and regulate the behavior of teachers. These attempts might be found in admonitory advice to teachers in daily newspapers or, during the 19th century, the growth of the popular novel. It was not until toward the end of the first half of the 20th century that the importance of interpersonal relationships in organizations emerged, from which it was possible to find some of the conceptual antecedents to modern theory and practice of personnel evaluation.

In 1659, Charles Hoole, an English grammar school master, published pamphlets that contained statements about teacher effectiveness, distribution of responsibilities between the master and his helpers (known as monitors), and the necessity for the teacher to maintain a favorable image with parents on whom his livelihood depended. It is interesting to note that Hoole struck a note that was to persist for almost the next century and a half in that country and others. It was the pupils themselves who were responsible for their learning, and any deficiencies could be attributed directly to them provided that the organization of the classroom by the teacher was competent. For instance, when Horace Mann visited schools in Massachusetts around the middle of the 19th century, he found that the pupils were regarded as being responsible for their own progress. Inability to learn was construed as laziness or lack of motivation.

Early in the 19th century the influence of the great headmaster of Rugby, Thomas Arnold, was strong in English public school circles. In fact, when the government-funded state grammar schools opened later in the same century, Arnold's Rugby became the example to emulate. Arnold stressed the importance of teachers maintaining their "Christian reputation" and being consistent in their behavior with pupils, but he also considered that the enforcement of strict discipline by teachers would have, as a natural consequence, a sound understanding by pupils of what was being taught.

The first coordinated, nationwide attempt to assess teachers, and reward them accordingly, occurred in England during the late Victorian era. This was called payment by results. Although it certainly took the mantle of responsibility for learning from the shoulders of the pupils, it would be difficult to imagine a more diabolical approach to education. Simply put, if pupils who attended government-funded boarding schools grasped prescribed basics of learning, then the teacher's meager income was augmented. The whole process was monitored by Her Majesty's Inspectors until 1902 when, as a result of public outcry, parliament brought to an end a practice that had corrupted education for two decades. In passing, it is worth questioning whether American school districts, which in the 1980s instituted

merit pay for master teachers, have fully considered the possible consequent narrowing of the focus of education.

The English inspectorial system has continued to the present day, both in the UK and many Commonwealth countries. Its functions have varied considerably over the years. Inspectors generally have acted as a regulatory authority ensuring that schools, by the nature of their adherence to public policy; are accountable for the expenditure of taxpayers' money. At least until the 1970s one specific function of the Inspector has been the assessment of teachers for promotion. Criteria for judgments have aligned themselves more with organizational requirements and conformation to written policies than the professional development of the teacher.

Into the 1900s

There are numerous indications in this country that, as one would expect, informal assessments of teachers by principals and parents were taking place from the beginning of this century. Some of these have been contained in principals' annual reports and those of school boards. There is no indication that formal, written procedures existed; but the public nonetheless developed a view of what a good teacher should be. Physical attributes, including personal grooming, as well as personal traits predominated as the criteria for judging the worth of the teacher. It was assumed that a teacher who spoke well, maintained a good appearance, and was enthusiastic, confident, and of sound integrity was a good teacher to whom students would respond by making pleasing progress. It was not until midway through the 20th century that it was realized that personality characteristics did not necessarily relate to the quality of teaching performance. It was even later before the truth was accepted that factors apart from the teacher influenced student achievement and that the effectiveness of a teacher in promoting learning is a most difficult concept to measure.

It is well known that research in education has trailed far behind that of industry but has often been influenced by it. It was assumed that concepts and methods developed for the organization of industry would work in schools and school districts. Charles Bobbitt (1912) was the most influential writer of the times; he tried to build connecting bridges between the theory and practice of industry and those of education. With the application of industrial techniques, particularly those of management, schools should produce predictable and improved results. These results should be linked specifically to society's requirements. Students were to be taught in such a way that society's expectations would be met. In other words, the students were the raw material of educational production. Some strands of what Bobbitt proposed, and many school districts tried to enforce, have lingered through to the present day. In this process teachers had to be utterly efficient and were judged

by their superiors by the extent to which designated goals of pupil learning were attained. In 1925 a National Education Association report stated that 75 percent of school systems in large cities were using various kinds of teacher efficiency ratings, a possible outcome of corresponding movements in industry. High among criteria were instructional techniques, personality, professional attitude, cooperation, and the maintenance of discipline records that incorporated classroom management.

The famous Hawthorne studies conducted by Mayo in the early 1930s and concluded the same decade by Roethlisberger and Dixon swept away adherence to the severely scientific approach to management and introduced, as a byproduct of research, the human relations era. Conforming to standardized expectations and plans of an organization gave way to the importance of interpersonal relationships and the concept that increased productivity should stem from this source. It may be possible to draw a historic parallel between the scientific and human relations dichotomy in management and approaches to teacher evaluation. One of the major problems that exists today when analyzing the true purpose of teacher evaluation is to decide whether the outcomes lead to a conforming with organizational standards and requirements or to teacher professional development based on effective interactions with students. The former gives emphasis to organizational growth while the latter increases student learning as a result of teacher development.

Post-World War II Until The Mid-1970s

During the 30 years under discussion, publications on the subject of the personnel appraisal function within public schools are replete with opinion-based literature but are lacking in research supported by empirical data. Nonetheless, both the literature and research that were carried out indicate that there was a growing consensus in some of the major aspects and purposes of teacher evaluation. Perhaps as an outcome of the Hawthorne studies, the futility of pursuing strategies that would lead only to ill feeling seemed largely to have been realized. This realization forced administrators to seek improved, more constructive appraisal methods. What appeared most important, perhaps, was the growing belief that the entire system must gain from improved teacher performance arising from more widely acceptable evaluation processes.

Bolton (1972) maintained that, whether or not formal appraisal processes take place, teachers are evaluated continuously.

They are evaluated by students, parents, other teachers, administrators, supervisors, and the public. The question is not whether teachers *should* be evaluated, since this cannot be avoided, but rather how systematic the evaluation should be in order to be most effective (p. 23).

He also considered that teachers have intrinsic desire to improve their performance. By contrast, Wolf (1971) contended that teachers are extremely reluctant to engage in evaluation exercises, although he readily admitted that there should be productive outcomes such as professional skill improvement, responsiveness to change, and accountability to constituencies that must be kept informed. On this and other salient aspects of teacher evaluation, opinions differ in the literature almost in proportion to its increasing proliferation.

Both the theory and practice of the teacher evaluation process related in this book indicate that modern day theorists and practitioners at least acknowledge the possibility of the appraisal function resulting in teacher development and, in many cases, see this as its prime objective. While the literature during the years leading up to the mid 1970s does not dwell on this aspect of teacher evaluation in any sustained fashion, it does relate to the possibility of positive outcomes either directly or indirectly.

A brief outline is given of the literature of the post-war period to the mid-1970s in these categories:

1. Systematic accountability of teachers
2. Teacher attitudes toward the appraisal process
3. Teacher learning as the basis for the appraisal function
4. Teacher competencies
5. Who appraises?
6. The relationship between teacher attitudes of personal ability and the appraisal function
7. The formative emphasis

1. Systematic Accountability of Teachers

During the 1960s and increasingly into the 1970s teacher evaluation attained growing importance. This was partly attributable to public demand for accountability in education which, by now, had shifted from a teacher's curriculum and program management to the quality of classroom teaching and student learning.

A national survey conducted by NEA in 1964 indicated that half the school systems followed formal procedures in the appraisal of their teachers and that written ratings were required in 3 out of 4 of the schools for probationary teachers and in 2 out of 3 for continuing teachers. Almost invariably, the principal was responsible for the evaluation process but occasionally shared that responsibility with other administrators. A few years later Stemnock (1969) not only found, as the 1964 survey had discovered, that principals are almost always responsible for appraisals, but he also was able to conclude that teachers strongly agreed that the

principal should be responsible for their professional accountability. Stemnock also found that 90 percent of the schools surveyed nationally had formal appraisal procedures of teachers. It is significant to note that in 1972, in another NEA survey, 55 percent of school systems had revised their teacher evaluation procedures during the previous 3 years.

That teachers consider themselves to be accountable for their professional conduct is strongly supported by the literature of this period. The survey conducted by Stemnock (1969) found that 90 percent of teacher respondents indicated approval of regular appraisals for professional accountability. Various writers, however, maintained that while teachers did not oppose accountability on professional grounds, they did strongly object to the form of accountability adopted by many school systems.

If teachers consider themselves accountable, or are considered accountable by others, it follows that the teacher evaluation process must have recognized purposes. Three studies showed large areas of agreement about the purposes of teacher evaluation.

Ingils (1970) analyzed samples of teacher appraisal programs from 70 school districts in 38 states. He discovered the following commonality of procedure and purpose:

1. To improve quality of instruction
2. To assist the teacher in areas that need improvement
3. To protect the competent teacher and eliminate the incompetent

Stemnock's (1969) investigation revealed that nearly 93 percent of responses from teachers favored undertaking evaluations for the purpose of assisting the teacher to improve competency. Interestingly, 54 percent of the responses also favored appraisals for the purpose of dismissing incompetent teachers. Only 17 percent, however, were in favor of using the process to determine advancement on the salary scale.

The NEA (1972) survey referred to earlier gave the following responses: 94 percent of teachers thought that evaluation should be used to stimulate improvement of teacher performance and 82 percent considered that evaluation should be used to establish evidence where dismissal from service is an issue.

All three surveys, therefore, clearly displayed a marked desire by school districts and individual teachers to give the highest priority to the improvement of teacher quality. Moreover, teachers felt accountable both to their profession and to their students.

2. Teacher Attitudes Toward the Appraisal Process

By now, the literature contained voluminous rhetoric about teacher attitudes to the evaluation process. For this reason, the only purpose of this section is to examine the dilemmas held by individual teachers about evaluation. Most importantly, research showed that teachers were willing to accept the principle of appraisal while at the same time rejecting methods adopted by their school or school system.

One of the dilemmas facing teachers then, and now, is the belief that, on the one hand, the evaluation function should lead to professional growth while, on the other hand, it provides a ready weapon for manipulation by administrators. What potentially should be good may be seen as functionally insidious. Gage (1973) provided a further dimension when he separated teacher optimism and administrative manipulation as aspects of teacher appraisal.

I have vague, private feelings that accountability reflects a fundamental struggle between those who possess some degree of trust in the developmental regularity of social and human organisms and those who trust only their own power to manage other people's comings and goings (p. 95).

A study in 1974 by Zelanak and Snider demonstrated that the perceptions of teachers about the evaluation process cannot be ignored. Their attitudes are important to the success of the process. The study compared the attitudes of teachers who believed the intention of appraisal was for administrative purposes with those of teachers who believed that the purpose of appraisal was aimed at improving instruction. This study strongly indicated that participating teachers who felt that the appraisals were means for instructional purposes were supportive of the process. By contrast, teachers who felt that appraisals were meant for administrative purposes—dismissal, tenure considerations, compilation of permanent record files, assignment modification, promotion—viewed the process in a very negative fashion.

Results of the Zelanak and Snider study are in general accord with views expressed often in the literature. If teachers are convinced that the evaluation process will reduce their status or in some manner act to their detriment in relation to their job function, it is logical that a negative reaction will result. It is equally reasonable to expect that teachers who are sincerely convinced that the principal's prime intention during evaluation is the improvement of instructional skills will be less intransigent to suggested changes in their approach to instruction.

Moreover, teacher doubts about the criteria to judge effective teaching, and to be used in the evaluation process, abound in the research literature. There was strong agreement that a tendency existed for the evaluator to focus on teacher traits and personal characteristics instead of behaviors directed at the effective management of learning conditions within the classroom and teaching skills themselves.

One obvious problem from the use of traits and characteristics, and one affecting teacher attitudes, is that it is highly improbable that any two evaluators could reach agreement on what it was that an effective teacher did when he or she was thought to possess particular traits. The implications are clear. An evaluator could make judgments about a teacher's performance on the basis of what he or she considered an effective teacher should be rather than on the basis of external standards whose credibility had been substantiated by behavioral meanings widely accepted by both the teacher and the evaluator.

By the early 1970s there was strong consensus among writers including Castetter (1971), Bolton (1972), and House (1973), and researchers such as Ryans (1960), Kleinman (1966) and Popham (1971) that teachers had little faith in either the ability or reliability of appraisal instrumentation. Investigations found that administrators most often judged teacher competence on the basis of (a) teaching ability, (b) disciplinary ability, (c) scholarship, and (d) personality, and demonstrated in their writing that teacher rating instruments and raters' assessments, commonly used in school systems, were unreliable. Studies concerned with the validity (whether content, correlational, or construct) of teacher evaluation instrumentation are both rare—apart from those dealing with student assessment of teacher competency—and inconclusive. The reason is not difficult to discover. The validity of an instrument depends upon the situation in which it is used; an instrument judged to be valid in one situation may be invalid if used in another situation for a different purpose.

3. Student Learning as the Basis for the Appraisal Function

In the history of teacher evaluation, there is no topic on which opinion varies so markedly as that of the validity of basing teacher effectiveness on student learning. Moreover, there is growing agreement today that there may be a nexus between particular teacher behaviors, based on effectively carrying out specified duties, and student learning. However, by the early 1970s the battle lines were drawn between those arguing against student learning as a basis for teacher evaluation and those supporting the contention.

Having made an extensive review of the research on the impact of teacher behaviors on student outcomes, Rosenshine and Furst (1971) concluded that there is little knowledge of the relationship between teacher behavior and student growth. Nonetheless, they did propose 11 teacher behavior variables affecting student learning that appear, from the perspective of previous research, to be the most promising of the variables studied to that time. In 1974 Heath and Nielson also summarized the findings of previous reviews of the research conducted on the relationships between teacher characteristics and student achievement over the past

50 years. These earlier reviews also generally concluded that educationally significant relationships had not been demonstrated, not because of minor flaws in the statistical analyses but more significantly because of sterile operational definitions of teaching and achievement.

Any discussion of student achievement brings in its wake a difficult criterion problem. This problem relates to the stability of various criteria and the reliability of their measurement. For instance, Glass (1974) criticized the use of standardized achievement tests to measure teacher effectiveness precisely because such tests do not reliably measure teacher effects on pupil gains in knowledge across a period of time. Moreover, Gage (1973) stressed that 1 of the major problems of previous competency research in teacher appraisal had been the overattention to a single criterion or, at the most, 2 or 3 criteria of effectiveness. This has resulted in the ignoring of many important classroom process variables; that is, the complete context in which teaching takes place.

Other writers and researchers stated unequivocally that teachers cannot be held responsible for student growth. They contend that knowledge about the processes of teaching and learning is so insubstantial that rational conclusions cannot easily be drawn. As an example, consensus had not been reached about basic skills. Nonetheless, appraisal programs had traditionally assumed a relationship between teacher behavior and educational outcomes. These writers did not deny that teachers may be appraised on the basis of competencies chosen for their validity as professional entities. What they did deny was that accountability may be based on educational outcomes that cannot be accurately measured.

Various writers considered it feasible for emphasis in the appraisal process to be placed on the development of teacher skills. For instance, Rosenshine and Furst (1971) listed five variables that show a strong relationship with measures of student achievement: clarity, variability, enthusiasm, task orientation, and student opportunity to learn.

Some, however, supported student learning as the basis for teacher evaluation. From their perspective, the assumption underlying teacher effectiveness is that the displaying of particular behaviors by a teacher results in particular student outcomes. This is unequivocally refuted these days by those who support Scriven's (1988) duties-based approach to teacher evaluation.

From about 1960, following the lead of industry, which had utilized performance objectives as the basis for judging personnel effectiveness, various educational researchers seeking solutions to the criterion problem had shifted from studying primarily what the teacher was doing (i.e., the means of instruction) to examining changes in learner behavior as a result of instruction (i.e., the outcomes of instruction). The criterion for evaluating teacher performance thereby became perceived change in student learning behavior.

Among leading writers in teacher evaluation who subscribed to this theory were McNeil (1967) and Popham (1971). In an attempt to isolate a valid indicator to evaluate teachers' instructional skills based on a measurement of students' attainment of instructional objectives, Popham developed a teaching performance test. Popham believed that the only important function of a classroom teacher is to promote beneficial changes in each learner. Each of his three teaching performance tests contained a set of specific instructional objectives measured by a posttest, the items of which vary between subject fields.

Researchers such as McNeil and Popham, in particular, have shown that by specifying changes in learners, arranging instructional events to produce the desired changes, and appraising the learners' attainment of instructional objectives, selective indices of teacher performance, based on student achievement, can be obtained. What Popham, McNeil, and others have done, however, is to place complete credibility of the teaching act upon one criterion. Researchers such as Rosenshine and Furst have strongly opposed such a view. Their contention is that no one criterion is complete and, moreover, a preference for one as opposed to another involves value judgment by the appraiser.

By their decisions, and actions, it was clear that many school districts and individual principals concerned with evaluation took the stance that there were cogent and valid competencies upon which teacher evaluations could be based. How extensive teacher and principal agreement was about these competencies was a problem not faced by research at that time. Without such agreement, it could be conjectured, the evaluation function could not readily result in development of teacher performance.

4. Teacher Competencies

By the 1970s the term "teacher competency" was thought to be any action taken by a teacher that contributes to the cognitive, affective, or motor-skill development of the student. According to this definition, emphasis is primarily placed on student growth.

The publication of a book by Ryans in 1960, *Characteristics of Teachers: Their Description, Comparison, and Appraisal*, was a landmark event in teacher evaluation. It was Ryans' contention that by identifying the characteristics of excellent teachers, it should be possible to use these attributes to undergird both teacher training and teacher evaluation. Ryans discovered that the development of this task was fraught with difficulties, mainly because very successful teachers often displayed quite different characteristics of effectiveness. Nonetheless, Ryans' work influenced significant studies by groups such as the National Center for Research

on Teacher Learning (Michigan State University) in their endeavor to discover the correlates of effective teaching.

Much of the consensus about the kinds of competencies an effective teacher should possess had been identified by the use of expert opinions of professional educators. On occasions, a factor analysis technique had been employed to bring to the surface the principal underlying concepts. Two of the more important teacher competency taxonomies, developed in this country from a process of extensive logical task analyses of teaching by principals, teachers, superintendents, and university educators, were the Houston Needs Assessment System (1973) and the Pennsylvania Competency Based Teacher Education Program (1971).

The Houston Needs Assessment System accepted, as its basic assumption, that effective teaching requires particular professional skills, attitudes, and knowledge. These were translated into objectives described in explicit, observable terms. Two premises underlay the Houston study: (1) that different teachers demonstrate varying levels of competencies and (2) that teachers as professionals are responsible for their own improvement. The taxonomy associated with the Pennsylvania program was designed as an instrument for teacher self-examination and development. It also anticipated that the competency inventory might serve the purposes of coordinating research into both appraisal criteria and procedures.

5. Who Appraises?

The significant research carried out in the 1960s by NEA (1964) and Sternock (1969) lent considerable weight to the contention that there seemed to be wide agreement among teachers that principals should have the responsibility for their evaluation. There was also wide agreement that teachers and principals must not necessarily reach consensus about appraisal criteria and practices if anything of a worthwhile nature was to result from the process.

Flanders (1970) proposed that teachers and administrators should institute competency contracts by which particular schools to be evaluated might be identified jointly. Baseline data on teacher performance could be gathered by some objective means mutually agreed upon. Performance criteria that represent school development in a particular direction could then be specified. Training and developmental materials to meet objectives would be made available to teachers. The final evaluation would be based upon attainment of a specified performance level.

If there was a growing consensus about the place of principals in the appraisal function, such was not the case about other possible evaluators. These include teaching peers, the teacher himself, and students.

The literature dealing with peer assessment by teachers is fragmentary and unsustained. Popham and McNeil expressed the opinion of many others when they

stated that teachers have traditionally been reluctant to make evaluation statements about the teaching skills and effectiveness of their peers. Teachers have expressed concern about the embarrassment that would arise if they failed in the eyes of their peers. It is interesting to note that, with the continued acceptance of teacher evaluation by teachers themselves and the professional nature of the procedures and stances adopted by their schools, teachers these days are considerably less reluctant to have their peers included in an evaluation panel.

On the surface, at least, it would appear that self-appraisal should reduce threat and increase the likelihood that the evaluation process will aid teacher development. It should enhance both involvement and acceptance of the process. While some writers and researchers viewed self-appraisal as a powerful means for a teacher to be a master of his own professional growth, Ryans (1960) contended that the chief disadvantage of self-appraisal is that the approach does not readily relate to outside criterion measures. Reported findings showed that there were at least two other disadvantages: teachers cannot accurately analyze specific aspects of their behavior because they lack a conceptual framework for observation, and teachers lack the technical competence necessary to operate such resources as video equipment to capture their behavior for analysis.

Whether or not students should evaluate teachers has been a vexing question. Historically, and realistically, most students have *always* assessed their teachers, at least informally. To this day some teachers contend that students, lacking the skills in training in instructional techniques and evaluation, should have no part in the process.

Nonetheless, during the seventies there was a growing body of research literature that began to change early skepticism to confidence in the ability of students—at least from intermediate grades to graduate school—to make reliable and valid judgments of teaching performance. Bryan (1959) maintained that student reaction reports could help teachers to determine the degree to which desirable characteristics exist, to discover unsuspected weaknesses and strengths, to obtain the proper balance and emphasis on competing factors in the teaching situation, and to get recognition for excellent teaching. One main criticism of student feedback that prevails to this day is that the halo effect inevitably distorts an otherwise balanced judgment.

6. The Relationship Between Teacher Attitudes of Personal Ability and the Appraisal Function

During this important formative period in the development of teacher evaluation, two well-organized studies showed that teachers' attitudes about the evaluation of

their teaching performance will strongly influence their ability to gain from the process.

Wagoner and O'Hanlon (1968) found that those teachers who hold favorable attitudes about appraisals are more likely to benefit than those who do not. Wolf (1971) found that teachers who perceive the evaluation process as important for decisions regarding both teacher and learning effectiveness tended to value student appraisals for decision making about teaching and learning. By contrast, teachers who perceived appraisals as important only for decisions related directly to student learning tended not to value students' judgments.

These studies invited conjecture concerning the influence of teacher attitudes toward the evaluation function despite sound and carefully chosen criteria and procedures. Certainly, the studies substantiated the opinions of those teachers who could not see evaluation serving their best interests but who were indeed willing to view evaluation as a vehicle to fulfill administrators' expectations.

7. The Formative Emphasis

The concepts of formative and summative evaluation emerged from the classic exchange between Cronbach (1963) and Scriven (1967). While summative evaluation involves developing conclusions about the merit and worth of a completed or stabilized process, formative evaluation consists of collecting and feeding back appropriate information for systematic and continuous revision of the ongoing process.

There is little doubt that appraisal of teacher performance had traditionally been of the summative type of evaluation. Such an appraisal is a final and, by inference, complete statement of a teacher's effectiveness and worth to the system. Clearly, this approach was one of the chief reasons for teacher discontent.

What Wolf, Bolten, House, and other writers advocated during the 1970s was formative teacher evaluation that would allow continuity of information, including feedback from principal to teacher, enabling a monitoring of the type and direction of teacher activities. It was strongly proposed that the process would afford a teacher with the potential for professional growth the opportunity to improve performance. Taken to its broadest limits, formative evaluation would allow a teacher to be evaluated for effectiveness and for his or her relationship to the school, or school district, context. It was also agreed that formative evaluation showed teachers *how they could change or develop*.

If teachers were able to engage in a more systematic appraisal process in which they would share in rule-making, their perceptions of the entire function undoubtedly would become more favorable. If this could happen and if teachers became more deeply involved in various aspects of their formative evaluation, then there

would be a reduction in threat. Even by the mid-1950s the importance of effective feedback as an integral part of formative evaluation was not stressed. Within the literature, very little emphasis was given to this aspect of the appraisal function; and when it did occur, differentiation between normative and nonnormative forms was seldom clarified. This shortcoming has certainly been readdressed in the literature from the mid-1970s until the present.

The Late 1970s To The Present

The most significant educational document to confront educators and the general public during this period was *A Nation At Risk*, published by the National Commission on Excellence in Education in 1983. Although the seeds of discontent about American education had been sown before the release of this publication and reform in many areas was imminent, its publication gave the American public a heightened awareness that reform in education, particularly at the elementary and secondary levels, was essential.

A great deal of *A Nation at Risk* centers on the need to improve teacher performance, the qualifications of those entering the profession, and retention of the best teachers. Almost overnight the movement toward increased accountability in education and a close scrutiny of its intentions and outcomes became a matter of national importance. Immediate outcomes have been a reassessment of teacher evaluation procedures by school districts, the realization that increased teacher competency will be a cornerstone to educational improvement throughout the nation, and the acceptance that considerably increased guidelines and standards to assess teacher evaluation systems are essential.

By 1983, 98 percent of school districts had some form of teacher evaluation model in use. These days it is most rare to find any school throughout the nation not practicing teacher evaluation. As educators search for ways to increase the effective approaches to teacher evaluation, they have discovered that the adoption of a process working well in another context may not be appropriate to theirs. For this reason university personnel skilled in the theory and practice of evaluation have increasingly been approached for guidance.

Formative and Summative Evaluations

One ever-present problem facing all school districts is the dilemma of choice between formative and summative types of evaluation. As this chapter has shown, this problem has prevailed at least since the 1950s.

Many school districts have adopted a predominately summative approach, with organizational aims and goals assuming greater importance than teacher development. Many other districts have adopted a formative, clinical supervision approach.

Very few instances can be found where formative evaluation is the sole type of evaluation approach. Many school districts have endeavored to incorporate elements of formative evaluation into their total process, which means, in effect, that an attempt is being made to meet the needs of both the organization and the individual through evaluation. While some school districts have proved that such a combination is possible, they have seen the increasing importance of thorough planning and organization and have endeavored to demonstrate to staff that personal commitment to the goals of the school district may be enhanced by the willingness of teachers to strengthen their teaching competency. Unless such a climate exists in the school district as an outcome of thorough discussion and agreement between all concerned parties, any attempt to make compatible summative and formative types of evaluation have failed.

Evaluation and Accountability

During the 1980s there was growing acceptance of school and teacher accountability. For all its faults and potential imperfections, teacher evaluation was seen as part of the educational process. Teacher evaluation was here to stay.

Educators and researchers are searching for more appropriate techniques of teacher evaluation to help substantiate the introduction of different and improved curricula. They are finding that it is impossible to review and assess the value of changed educational programs without a close scrutiny of teacher attitudes and performance and, closely allied to these processes, student learning.

Occasions will arise where public demands for accountability alter what the school itself is endeavoring to do. Many school districts, however, have found that teacher evaluation has led to more productive working relationships within the school and district, wider understanding of the educational context by teachers, and a professional desire to improve personal skills. As these school improvements become more obvious, and the public generally realizes that educators know that they are accountable, the nexus between teacher evaluations and accountability is established.

In 1993-94, the Office of Educational Research and Improvement, the National Center for Education Statistics, Westat, Inc., and CREATE collaborated to conduct a national survey of public school teachers of kindergarten through grade 6 (National Center for Education Statistics, March 1994). Most teachers reported that their evaluations accurately reflect their teaching performance and are useful for

improving teaching. The teachers were in substantial agreement on the following points:

1. The practice of evaluating elementary school teachers is well established in their schools.
2. Their evaluations are guided by written policies.
3. Teachers are informed beforehand of the criteria to be applied in evaluating their performance.
4. Most teachers are evaluated by their school principal.
5. The main evaluation method employed is classroom observation.
6. Teachers receive both written and verbal feedback following their evaluation.
7. Teachers are given the opportunity to append their written response to the evaluation and/or to file an appeal.
8. Teachers are supportive of evaluations employed to improve their teaching skills.
9. Teachers view uses of evaluations to discharge incompetent teachers or, especially, to award merit pay as less important than uses of evaluation to improve teaching.
10. Nevertheless, teachers indicate that evaluations should be used more than presently is the case to terminate incompetent teachers and determine teachers' pay levels.

The greatest percentage of the teachers noted that evaluations of their performance should consider overall teaching performance, subject matter knowledge, classroom management, instructional techniques, helping students achieve, and unique teaching demands. However, a much smaller percentage reported that those aspects of teaching were actually considered to a great extent in their last evaluation.

This discrepancy should receive attention in efforts to improve evaluations of elementary school teachers. Clearly, the dominant practice of evaluating teaching mainly or only on the basis of classroom observations is not a sufficient means of evaluating the full range of important teaching responsibilities.

Teacher Certification

A Nation At Risk makes clear that states and school districts must attract highly qualified and worthy people into the teaching profession. The diminution in this regard was most marked during the 1970s. The same publication stated unequivocally that incompetent teachers should be evicted from the profession. Although most Americans would not disagree with this contention, the legal mechanisms for

terminating the appointment of a poor teacher are most difficult. Moreover, most other professions and industries are more financially attractive than teaching.

The concept of certification of teachers on a statewide basis has been practiced in this country, and other places such as Australia, for many years to ensure that all applicants for the various stages of certification [according to years of experience] have basic professional and academic qualifications and that they reach satisfactory levels of competence, according to statewide criteria for judgment, during their probationary period. All Australian and U.S. states have adopted a basically similar approach.

In the United States, the Southern Regional Education Board has recommended that the complexity of certification be reduced so that states are able to move to a common certification test and that the graduate courses that teachers take for recertification relate to teaching assignments. The National Commission on Excellence in Education and the National Science Foundation both made a particular point of recommending the exploration of ways to allow outside experts to teach (as for example, guest lecturers, experts-in-residence) as part of teaching teams. Some alternative certification programs have been established in order to recruit highly educated persons, with no background or preparation in teaching, into hard-to-fill teaching positions, e.g., Teachers for Chicago.

Most states have already adopted teacher competency tests, such as the National Teacher Examinations, for teacher certification. Others are again establishing professional standards and practices boards. Some, like Oklahoma and Michigan, are working to strengthen the teaching profession by adopting higher admissions standards for colleges of education together with competency tests for certification and recertification, teacher evaluation, and forms of continuing teacher education through inservice and other means. At present the Carnegie Foundation is undertaking valuable work in the area of national certification examinations for those wishing to be recognized for excellence in teaching. Also, Educational Testing Service replaced its National Teacher Examination (NTE) with its new Praxis series.

It is interesting to note that during the 1980s the huge proliferation of literature and media comment about education and teacher evaluation increasingly included statements and opinions from those not closely connected with the formal educational process. This is true worldwide. There appears to be a common thread running through much of this comment that adds an onus of responsibility to teachers. Teachers are advised to be aware of the exact outcomes of their efforts by analyzing the quality of student learning. They should teach in different ways using a variety of media to suit different purposes, student needs, and social problems. Teaching must be well planned, clearly executed, and supported by student feedback. A whole range of higher order skills must be understood by teachers and related to the learning environment. The increased weight of responsibility on the

teacher, and on the principal who will have to evaluate the teacher's performance, is most daunting, to say the least.

Educational Testing Service recently completed a 6-year study to create assessments for licensing beginning teachers in developing The Praxis Series: Professional Assessments for Beginning Teachers. This effort resulted in the formation of classroom performance assessments. As the project developed, a methodology was created to address the problem of defining good teaching in a way appropriate for assessment purposes while remaining true to teaching as experienced by expert practitioners. Carol Anne Dwyer, a leader in this project, gives a detailed account of criteria for performance-based teacher assessments in Chapter 2.2 of this book.

In October 1987, the National Board for Professional Teaching Standards began its work to develop techniques, based on rigorous standards, to recognize the knowledge and skills of experienced, outstanding teachers. It is the intention of the National Board to act as a catalyst to improve schools through the national certification of these accomplished teachers. This project, which should be fully operationalized by 1997, is described in Chapter 4.8 of this book.

Around 1980, 2 statisticians, William Sanders and Robert McLean of the University of Tennessee, began to explore the feasibility of using statistical mixed-model methodology to overcome many of the existing impediments for placing student achievement data in an outcome-based assessment system. This developed into what is now known as the Tennessee Value-Added Assessment System (TVAAS), which assesses the impact of educational systems, schools, and teachers on student gains, comparing performance measures over at least 2 years using norm-referenced achievement tests. It is the aim of TVAAS to provide unbiased measures of student academic progress and thus to provide direction for strengthening the educational policies and programs of the state, districts, schools, and teachers. The strength of this approach is its huge statewide, longitudinal schools database supported by a powerful computer system, its employment of powerful statistical procedures, its grounding in systematic tests of its underlying statistical assumptions, its careful and unrushed development, and its strong, sustained efforts to inform and involve stakeholders. The project's main limitation, to this point, has been its reliance on multiple choice, norm-referenced tests. This model is explained in some detail by William Sanders and Patricia Horn in Chapter 5.8 of this book.

Along side the Tennessee project, there have been 2 parallel projects devoted to directly using measures of student growth to evaluate teaching effectiveness. The Dallas Independent School District, under the leadership of William Webster, uses a range of student performance indicators to assess teaching effectiveness at the school level. A districtwide accountability commission defines the student outcome measures, which extend far beyond multiple choice achievement tests (that, to this point, Tennessee has used as its exclusive measure of student growth). Webster and his colleagues employ multiple regression techniques to partial out a school's

unique effects on measured student growth from a wide range of student background and other influential context variables. Dallas will soon use its student outcomes system to evaluate the effectiveness of individual teachers. Webster describes the Dallas system in Chapter 4.10.

Oregon is developing yet a third approach to evaluating teachers based on assessed learning gains of their students. Dr. Del Schalock and his colleagues at Western Oregon State College are leading this effort. Under this approach a teacher is required to produce evidence of student growth for 1 or more designated units of instruction. The teacher develops parallel forms of a work sample performance assessment exercise keyed to the objectives of the instructional unit. The teacher administers the work sample exercise at the outset of the instructional unit and at its end. The teacher then scores the pretests and posttests for each student and computes an index of growth. Then external evaluators evaluate the results produced by different teachers in order to separate those who are showing acceptable levels of student gains from those who are not. At present, Oregon is using this approach mainly in an experimental mode, but is considering adopting it as part of the assessment used to certify teachers. The Oregon researchers need to solve some significant problems with this approach. Teachers are producing substandard work samples. Often the posttest and pretest exercises are the same. Moreover, the most work samples are keyed to low-level objectives. They produce unreliable measures, and they are not standardized across teachers. The virtues of the approach are that it engages teachers in development of performance assessments keyed to instructional objectives and focuses the teachers' attention on student learning. In its present state of development, the approach appears to have little or no merit to support high stakes decisions about teachers. Its main value is in helping teachers to learn about and integrate performance assessment into instruction.

The Legal and Political Aspects of Evaluation

With the greatly increased interest in teacher evaluation, and its practice in schools, there has been a flurry of legal and political activities regulating and monitoring evaluation processes.

Almost without exception, school districts have policies governing the evaluation of teachers. These policies range from simple statements affirming the district's obligation to ensure quality instruction to elaborate documents addressing a wide range of evaluation issues. As has been mentioned, more often than not they will reflect the district's personnel management priorities and the teachers union's negotiated conditions for cooperating in the evaluation process. Policies usually address areas such as the purpose of evaluation, objectives to be obtained and standards for a satisfactory level of performance, frequency of evaluation and

personnel involved, remediation procedures for poor performance, detailed procedures of the evaluation process, and, increasingly, provision for professional development. Good policies are the foundation for sound, and acceptable evaluation processes. They usually are supported by state legislation.

By 1984, 46 states had a law or administrative regulations mandating the evaluation of teachers, and by 1994 only 1 state had not completed such a mandate (but plans to do so shortly). The predominant number of these states included professional improvement of teachers as a purpose of evaluation. Influenced by a number of forces, these mandates typically were designed to protect the public from incompetent and unethical educational processes while aiming also to preserve the due process rights of teachers.

Although the regulations and the various state acts are often dissimilar, most have endeavored to address areas like performance standards, forms and procedures relevant to the evaluation process, the timeliness and formalities that must be observed in report writing, and grounds for dismissal and procedures for appeals.

There are inconsistencies among the various state laws related to teacher evaluation. Some states specify a single method for data collection, while others suggest multiple approaches. Ten states mandate classroom observations, 9 require interviews, and 6 make a provision for the review of work portfolios. It should be noted that the various state mandates give minimal acceptable standards for teacher evaluation practices. By statement or implication, school districts are free to issue more detailed, varied, or stricter procedures.

Unions have closely monitored the introduction, implementation, and processes of teacher evaluation schemes. Collective bargaining agreements have invariably been associated with the endeavors of school districts to introduce teacher evaluation. During the 1980s, in particular, contracts have specified the purposes of evaluation, methods of information gathering, frequency and conduct of classroom observations, processes for reporting of results, teacher involvement in the process, and appeal procedures in the case of adversely critical remarks or notification of dismissal.

While it is possible for collective bargaining to skew teacher evaluations toward a regulatory process rather than teacher improvement, the predominant number of teacher unions have accepted evaluation as an essential adjunct to professional improvement. Most have insisted upon at least some formative or improvement-orientated aspects being included in agreements. Supported by state legislation, bargaining agreements invariably build in due process safeguards.

Agreements also are likely to contain minimally acceptable teaching standards so that in the case of subsequent court action it may be possible to specify how a teacher's performance violates those standards. Beckham (1981) recommends that to face judicial scrutiny, an evaluation policy must include (a) a predetermined standard of teacher knowledge, competencies, and schools; (b) an evaluation

system capable of detecting and preventing teacher incompetency; and (c) a system for informing teachers of the required standards and according them an opportunity to correct teaching deficiencies.

Future Challenges In Evaluation

It is all too easy for critics of current evaluation schemes to complain that the process is fraught with difficulties. So it is. On the other hand both the public generally and educators themselves understand the importance of teacher evaluation and its essential part of the educational process. Moreover, many teacher evaluation schemes presently being practiced in this country undoubtedly are serving the important function of making assessments of teacher qualifications more objective than they otherwise would be. Practitioners are more aware than anyone of the imperfections of their particular approaches and obviously wish to see improvements take place. Such improvements will arise only from a standards-based analysis of schemes *as they exist* and not by the carping criticisms of theorists offering advice from the sidelines.

The challenges to be faced include the questions of how closely evaluation processes should be associated with the question of merit pay for master teachers and who should be involved in the evaluation process (including, perhaps, parents), a closer analysis of the actual styles and approaches to teaching—whether it is to be classified as labor, or craft, or profession, or art—and the influence on teaching of various contexts and social environments (see Wise & Darling-Hammond, 1985). In summary, while there is total agreement that teachers must engage their students in active learning in interesting and imaginative ways, the exact definition of what that “something” is will continue to tax the minds of educators and indeed all concerned with education.

An experienced, perceptive principal, nonetheless, will discern a teacher's competencies and fulfillment of prescribed duties and their appropriateness to effective student learning. The challenge to the art of evaluation is to define and assess more closely each teacher's responsibilities so that teacher evaluations become more fair to the individual and useful for school improvement. Any astute evaluator is fully aware of the fact that there is no such thing as uniform teaching behaviors nor that learning results only from what is occurring in the lesson being observed. Researchers and writers of the 1980s such as Centra and Potter (1980) have observed that a teacher's influence may be small when compared with the totality of the effects of other influences affecting student learning.

The main challenge facing those concerned with evaluation is the purpose of the appraisal process itself and the desired outcomes. The problem that has dogged

both theoreticians and practitioners is how the process can both achieve organizational ends and increase the skills and self-esteem of teachers.

If the goals of teacher evaluation are decided by external authorities and by behavioral objectives and anticipated outcomes set by them, the evaluation will be summative, as we have seen. Policy and procedures will include definition of staff roles, set rules of procedure, specification of aims, and modes of official actions. During the course of the evaluation, scant concern will be given to improving the teacher's competence and performance being evaluated. The organizational context may be largely ignored or taken for granted. Awards and sanctions will either be implied or written into procedures.

By contrast, the professional development style of evaluation will involve the teacher concerned in all aspects of its planning and will impose a minimum of bureaucratic procedures. Feedback, reformation of goals, and positive encouragement will aim toward teacher improvement. Teacher autonomy will be viewed as both worthwhile and productive.

As has been mentioned, many school districts endeavor to incorporate both formative and summative elements in their evaluation schemes, with the latter predominating. Close documentation of these emerging schemes and practices should be made to determine whether, and under what circumstances, the two elements can be incorporated into a single approach. History may record that they are incompatible unless they are controlled and administered separately.

Also, it may prove more productive to design systems not for their formative or summative orientations but in terms of the decisions that need to be informed at different levels, e.g., the teacher, the principal, the district, and the state. Then the needed system(s) of evaluative information could be designed expressly to provide defensible feedback to serve decision making at each of these levels.

Standards for the Assessment of Systems for Evaluating Teachers

No account of the historical perspectives of teacher evaluation would be complete without comment on a recently completed, nationwide attempt to formulate standards for planning and implementing teacher assessment systems. Developed by the Joint Committee on Standards for Educational Evaluation, *The Personnel Evaluation Standards: How to Assess Systems for Evaluating Educators* (1988) is beginning to have a significant influence on teacher evaluation. For example, in 1994 Texas adopted an adaptation of *The Personnel Evaluation Standards* as state policy for teacher evaluations. Also, the *Standards* were referenced in four 1994 court decisions in Michigan concerned with demands by members of the public that evaluation reports of named teachers be released for public review. The *Standards* explicitly require that personnel evaluation reports be released only to

users and for uses that were previously authorized. The *Standards* warn that releases and uses determined after the evaluation has been conducted harm the public interest—by motivating supervisors and teachers to be very guarded, even superficial, in what they consider and report in evaluations—and violate the teacher's right to due process in making and honoring decisions that only specified, right-to-know audiences will see the evaluation findings and that they will use the findings only for the prespecified purposes. After studying Part 3 of this book, which deals with the *Standards*, readers may judge for themselves the extent and degree to which the *Standards* will influence teacher evaluation in the years to come.

Although writers and researchers by statement or implication have considered it a necessity to have standards for personnel evaluation, the Joint Committee's publication is the first systematic and detailed attempt to achieve this most difficult task. Public hearings and trials conducted throughout the nation indicated that the *Standards* will receive widespread endorsement and use. The product of a collaborative effort by numerous interested professional associations and their members, the *Standards* have undergone extensive review and refinement. Nonetheless, they are subject to further review and revision. In this respect they should not differ from any teacher evaluation system being used today. Moreover, it is hoped that the *Standards* will encourage all interested in improving education to devise better systems for evaluating teachers.

Attempts to Introduce Standards. In reality, until the present, an attempt has never been made to state, let alone develop, standards for teacher evaluation. Various attempts have been made to introduce standards for student performance and, on occasions, it was assumed that by some osmotic act these would indicate standards of teaching performance. A brief attempt will be made to look at some of the significant thrusts in educational evaluation in the United States.

The first major movement, which commenced in the early part of the 20th century and quickly gained momentum, was concerned with assessment of student performance; this was embodied primarily in the standardized testing movement. A half century later a second movement involved the evaluation of projects, especially externally funded projects. The third concerned evaluation of teachers and other educators in the manner outlined in this chapter. Although educational evaluation has become important since the 1970s, serious thought was not given to parallel standards until late in that decade.

Standards in the area of student evaluation appeared first in the 1950s in the form of the NEA publication, *Standards For Educational and Psychological Tests*. Standards relating to the evaluation of curriculum project areas were published in 1981 by the Joint Committee on Standards for Education Evaluation. This was entitled *Standards for Evaluations of Educational Programs, Projects, and Materials*.

In brief, then, it appears that efforts in standard setting moved from a focus on student performance, as reflected in the use of standardized tests, to the evaluation of programs as one possible crucial reason for deficiencies in students' performance, to standards regulating the systems for evaluating the main actor in education, the teacher.

Events followed this pattern. In the middle 1960s massive federal efforts were undertaken to improve programs for disadvantaged children, with a strong requirement that all programs must be evaluated. Nonetheless, the program evaluation of the late 1960s and early 1970s largely excluded concern for the evaluation of teachers who were undertaking the programs and projects. Convenient assumptions were raised that program deficiencies were due not to the persons involved, but rather to the concepts and designs of the programs themselves. There was a strong reticence to make individual teachers personally accountable for shortcomings of these programs.

As the review of literature from the important research and writings of the 1970s indicates, pressure increased dramatically to make those involved in education accountable for program quality. This is evident in the two teacher evaluation handbooks produced by the National Council on Measurement in Education (Millman, 1981; Millman & Darling-Hammond, 1991) and in Rand Corporation reports (Wise & Darling-Hammond, 1985; Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984). Teacher evaluation systems developed and proliferated without themselves being assessed by any set of standards. While many argued that state legislation in the late '70s and early '80s constituted a formal standardized approach to teacher evaluation, the inconsistencies and hasty development in the early stages fell far short of a cohesive and acceptable set of standards.

As a consequence the professional societies in education increased their efforts to develop sound personal evaluation and, as one measure, 14 of them supported the Joint Committee to develop the standards that were outlined in Chapter 2.

The Joint Committee Standards. The development of these Standards provides a vital step toward helping the profession not only to improve personnel evaluation, but also to integrate that work effectively with other forms of evaluation, particularly of student needs and performance, program plans, operations, and outcomes.

Beginning in 1985 a large number of people, representing different professional perspectives contributed to defining shared principles for both guiding and assessing personnel evaluation, work, and education. After initial planning, a national panel of writers was chosen largely through nominations by the sponsoring organizations. These produced alternative versions of a suggested list of standards. The first draft was scrutinized closely by both the national and international review panels in 1986 for changes, improvements, and developments.

Following a critique from an independent validation panel, field tests and national hearings of the revised draft *Standards* took place. The Joint Committee met in July 1987 to make decisions for finalizing and publishing the *Standards* and for dissemination planning. Subsequently, the Committee promoted the *Standards* and set in motion the process of periodic review and revision.

These *Standards* invite educational institutions of all kinds to recognize a long-standing need to have a sound evaluation process for entry into professional training, certifying competence, defining roles within institutions, selecting job applicants, monitoring and providing feedback about performance, counseling for staff development, determining merit awards, and making decisions about tenure, promotion, termination, and state or national recognition.

The *Standards* are aimed at improving present systems and practices. Their potential appears very significant in the history of teacher evaluation. At this writing, the state of Texas has adapted and formally adopted the Personnel Evaluation Standards to serve as the basis for reforming its statewide teacher evaluation system.

A National Center for Research on Teacher Evaluation and Dissemination of Outcomes

Funded by the U.S. Department of Education through its Office of Educational Research and Improvement (OERI), the Center for Research on Educational Accountability and Teacher Evaluation (CREATE) commenced work in 1990 at The Evaluation Center, Western Michigan University, which also houses the Joint Committee on Standards for Educational Evaluation. CREATE, which completes its funding cycle in 1995, has a mandate to address evaluation issues regarding the school as a whole and professional personnel employed to serve the school and school district. Of the five programs, one focuses on improvement of teacher evaluation. Combined, the five programs are designed to serve the nation's school evaluation needs by obtaining and synthesizing available knowledge on personnel and school evaluation, developing new knowledge and evaluation tools, and disseminating research findings and products for use in the nation's schools.

This book itself is partly an outcome of CREATE.

CREATE devotes primary attention to teacher evaluation. Five projects comprise the program entitled **Improvement of Teacher Performance Evaluation**. These address evaluations used to select and review the performance of teachers. One, the **Teacher Evaluation Models Project (TEMP)** has identified and assessed the strengths and weaknesses of extant teacher evaluation models and has developed an extensive list of teaching duties. A second, the **Improved Teacher**

Evaluation Models Development Project, is currently developing a duties-based model of teacher evaluation for implementation and testing in the Dallas Independent School District.

The third project, **The Evaluation Theory Development Project**, through study of school-based teacher evaluation systems, developed a theoretical framework and practitioners' GUIDE (see Chapter 3.2 of this book) for assessing present teacher evaluation programs. A fourth project, the **Teacher Self-Assessment Project**, gives emphasis to collecting, synthesizing, and testing evaluation strategies that teachers may use to assess and improve their personal instruction and classroom learning environment. The remaining project, the **Expert Science Teacher Evaluation Model**, was developed to respond to key questions: What is good science teaching? How can we recognize it when we see it? How can we evaluate it?

Through its **National Evaluation Resources Service CREATE** disseminates evaluation information and products to educators and policymakers nationally. The **Dissemination Program** engages in a wide range of activities, including the CREATE newsletter, *Evaluation Perspectives*, and the annual National Evaluation Institutes, which bring together practitioners and researchers interested in developing educational evaluation. The Institutes stimulate interest in national networking in teacher evaluation and other aspects of educational evaluation.

In its short life span, CREATE has already made a significant impact, and there is promise of further developments.

The Praxis Series: Professional Assessments for Beginning Teachers

The Educational Testing Service (ETS) has long offered to state education departments and school districts the National Teachers Examination (NTE) as a device for certifying that beginning teachers possess an acceptable level of content and pedagogical knowledge. In 1992, ETS replaced NTE with Praxis (Dwyer, 1992; 1993; 1994). This new series goes beyond testing of teachers and includes a set of classroom performance assessments. These are grounded in the constructivist theory of teaching and learning and include criteria that authoritative groups believe constitute the knowledge and skills needed to perform effectively in a given teaching domain in a variety of teaching contexts. The Praxis III Classroom Performance Assessment Criteria are in 4 groups: organizing content knowledge for student learning, creating an environment for student learning, teaching for student learning, and teacher professionalism. It is interesting that these criterial categories, which were developed through a consensus process, are quite compat-

ible with the system of duties defined by Michael Scriven (1994) from a philosophical standpoint. His categories of duties include subject-matter knowledge, instructional skill, assessment skill, professionalism, and "other duties." It seems that agreement is growing that teachers must be assessed on the effectiveness of what they do in classrooms rather than on how they do it. Thus, there seems to be growing consensus that effective performance of duties rather than proper use of preferred styles of teaching form the appropriate basis for judging teaching.

The National Board for Professional Teaching Standards

Another recent and noteworthy development in teacher evaluation is the work of the National Board for Professional Teaching Standards (NBPTS). This Board was created in 1987 pursuant to a recommendation of the Carnegie Task Force on Teaching as a Profession. In its 1986 *A Nation Prepared: Teachers for the 21st Century*, this task force concluded that improving teacher standards is the key to improving school effectiveness and that the status of the teaching profession must be raised substantially in order to attract and retain excellent teachers. To address this national need, NBPTS started work in 1987 toward transforming the teaching profession by establishing high and rigorous standards for what teachers should know and be able to do and setting up a network of assessment laboratories through which teachers could have their competence assessed and confirmed as worthy of receiving a national certificate of excellent teaching competence. With funding of about \$25,000,000 per year from the U.S. Congress, NBPTS has moved systematically to define standards and develop assessment devices and methods for a wide range of teaching content and grade level areas. In general, an applicant teacher is informed of what standards apply to her or his field of teaching, then develops a portfolio of information to document teaching effectiveness in her or his school, and subsequently completes a range of assessment exercises at an NBPTS assessment center. NBPTS then examines the portfolio materials and assessment center results in order to determine whether the teacher has met the Board's criteria and, if so, awards the national certificate of excellent teaching competence.

Of course, the NBPTS evaluations must themselves be evaluated to assure that they are fair and lead to justifiable certification decisions. Accordingly, NBPTS established criteria for evaluating the assessments and set up an independent laboratory at the University of North Carolina, Greensboro to apply the criteria in assessing the work and products of each assessment development laboratory. The criteria for NBPTS assessment systems include Administrative feasibility, Public acceptability, Professional acceptability, Legal defensibility, and Economic affordability (APPLE) (Baratz-Snowden, 1991, p. 145). Nyirenda (1994) has de-

scribed the metaevaluation system being used by NBPTS to evaluate NBPTS evaluations.

The NBPTS aims are responsive to critical needs for better teaching and learning in the nation's schools, the NBPTS budget is enormous, and the NBPTS work is systematic and highly participatory. This is a bold and expensive effort. It is too early to know if the NBPTS system will raise the standards of teaching throughout the U.S. and especially in schools in economically disadvantaged neighborhoods. It will be critical for NBPTS to conduct and use evaluation to help assess accomplishments and detect and correct problems. It will be important for the education profession and the U.S. society to evaluate the success of this effort as well, so that they can use and continue to support the work if it returns good value to the society or, if it is found to have poor cost-effectiveness, to replace it with something that works better.

Conclusion

The history of teacher evaluation, particularly before 1970, is a difficult and necessarily imprecise undertaking. As we have seen, formal evaluation systems have only recently emerged and are still beset with doubts about their efficacy. Nonetheless, as this chapter indicates, there are sufficient imperatives existing to ensure that teacher evaluation is now very much a part of the broad concept of education itself.

Although history all too clearly shows the deficiencies of various schemes to evaluate teachers, it nonetheless shows that valiant and increasingly coordinated attempts are being made to develop its acceptance and professional credibility.

Such are the calls for accountability in education, from the public generally and from teachers themselves to increase their professional standing, that teacher evaluation processes must improve. The fact that there are weak elements present is not an argument against teacher evaluation; rather, it is an argument for using better evaluation procedures, particularly those that focus on performance and those that have been subjected to empirical development.

Moreover, effective programs to improve teacher evaluation practices will need to build on lessons from the past. In this chapter, we have noted the following:

1. Traditionally, U.S. teachers have not been held in high esteem; society through its press for more rigorous and consequential teacher evaluation has denoted that teacher competence and professionalism are suspect.

2. Teachers have gained power through collective bargaining, resulting in many places in gridlock over teacher evaluation between the school authorities and the teachers union.
3. The attempt to improve teacher evaluation by finding the research-based indicators of effective teaching, for a time, carried an aura of scientific respectability, but subsequently failed and became discredited.
4. There remains a persistent quest to find defensible ways to assess teaching effectiveness based on student learning gains.
5. There also is a renewed interest in directly assessing teacher performance of assigned duties.
6. There is a growing consensus that whatever evaluation approach is used, it must help teachers to improve teaching competence, performance, and effectiveness.
7. There are as yet no clear winners among the competing approaches to teacher evaluation.
8. *The Personnel Evaluation Standards* provide a solid foundation for guiding and assessing the further efforts to improve teacher evaluation.

The chapters contained in this book outline promising theory and practice. While all theorists and practitioners referred to in these pages willingly admit that their ideas and procedures are open to scrutiny and improvement, they are firmly of the opinion that teacher evaluation is a vital component of progressive schools and school districts.

References

- Baratz-Snowden, J. (1991). Performance assessments for identifying excellent teachers: The National Board for Professional Teaching Standards charts its research and development course. *Journal of Personnel Evaluation in Education*, 5(2), 133-145.
- Beckham, J. C. (1981). *Legal aspects of teacher evaluation*. Topeka, KS: National Organization on Legal Problems in Education.
- Bobbitt, J. F. (1912). The elimination of waste in education. *Elementary School Teacher*, 12, 260.
- Bolton, D. L. (1972). *Selection and evaluation of teachers*. Berkley: McCutchen.
- Castetter, W. B. (1971). *The personal function in educational administration*. New York: MacMillan.

- Center for Research on Educational Accountability and Teacher Evaluation (1994). *CREATE contributions*. Kalamazoo: The Evaluation Center, Western Michigan University.
- Centra, J. A. & Potter, D. A. (1980). School and teacher effects: An interrelational model. *Review of Educational Research*, 50(2), 273-291.
- Cronbach, L. J. (1963). Course improvement through evaluation. *Teachers College Record*, pp. 672-83.
- Dwyer, C. A. (1993). *Development of the knowledge base for the Praxis III: Classroom performance assessment criteria*. Princeton, NJ: Educational Testing Service.
- Flanders, N. A. (1970). *Analyzing teacher behavior*. Reading, MA: Addison-Wesley.
- Gage, N. L. (Ed.). (1973). *Mandated evaluation of educators: A conference on California's Stull Act*. Washington, DC: Educational Resources Division Capital Publications, Inc.
- Glass, G. V. (1974). Teacher effectiveness. In H. L. Walberg (Ed.), *Evaluating educational performance*. Berkley: McCutchen.
- Heath, R. W., & Nielson, M. A. (1974). The research basis for performance-based teacher evaluation. *Review of Educational Research*, 44, 463-484.
- Hoole, C. (1868). Scholastic discipline. *American Journal of Education*, 17, pp. 293-324.
- House, E. R. (1973). *School evaluation: The politics and process*. Berkley: McCutchen.
- Ingils, C. R. (1970). Let's do away with teacher evaluation. *The Clearing House*, 44, 451-456.
- Joint Committee on Standards for Educational Evaluation. (1981). *Standards for evaluations of educational programs, projects, and materials*. New York: McGraw Hill.
- Joint Committee on Standards for Educational Evaluation. (1988). *The personnel evaluation standards: How to assess systems for evaluating educators*. Newbury Park, CA: Sage.
- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards, 2nd edition*. Thousand Oaks, CA: Sage.
- Kleinman, G. S. (1966). Assessing teacher effectiveness: the state of the art. *Science Education*, 50, 234-238.
- McLean, R. A., & Jordan, W. L. (1984). *Objective components of teacher evaluation—a feasibility study*, (working paper no. 199). Knoxville, TN: University of Tennessee, College of Business Administration.
- McNeil, J. D. (1967). Concomitants of using behavioral objectives in the assessment of teacher effectiveness. *Journal of Experimental Education*, 36, 69-74.

- Millman, J. (Ed.). (1981). *Handbook of teacher evaluation*. Beverly Hills, CA: Sage.
- Millman, J., & Darling-Hammond, L. (Eds.). (1990). *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers*. Newbury Park, CA: Sage.
- National Board for Professional Teaching Standards, 3rd Ed. (1991). *Toward high and rigorous standards for the teaching profession*. NBPTS: Detroit, MI and Washington, DC.
- National Center for Education Statistics. (1994). Statistical analysis report: *Public Elementary Teachers' Views on Teacher Performance Evaluations*, Fast Response Survey System. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, NCES 94-097.
- National Commission on Excellence in Education (1983). *A nation at risk: the imperative of educational reform*. Washington, DC: U.S. Government Printing Office, No. 065-000-00177-2.
- National Education Association. (1955). *Standards for educational and psychological tests*. Washington, DC: Author.
- National Education Association, Research Division. (1964). *Evaluation of classroom teachers*. Research Report, 1964-R14, Washington, DC.
- National Education Association, Research Division. (1972). *Evaluating teacher performance*. ERS Circular No. 2, Washington, DC.
- Nyirenda, S. (October, 1994). Assessing highly accomplished teaching: Developing a metaevaluation criteria framework for performance-assessment systems for national certification of teachers. *Journal of Personnel Evaluation in Education*, 8(3), 313-328.
- Popham, J. W. (1971). Performance test of teaching proficiency: Rationale, development and validation. *American Educational Research Journal*, 52, 105-117, 559-602.
- Rosenshine, B., & Furst, N. (1971). Research on teacher performance criteria. In B. O. Smith (Ed.), *Research in teacher education: A symposium*. Englewood Cliffs, NJ: Prentice-Hall.
- Ryans, D. G. (1960). *Characteristics of teachers: Their description, comparison and appraisal*. Washington, DC: American Council on Education.
- Scriven, M. (1967). The methodology of evaluation. *American Educational Research Association monograph series on curriculum evaluation*, 1, 39-83. Chicago: Rand McNally.
- Scriven, M. (1988). Duty-based teacher evaluation. *Journal of Personnel Evaluation in Education*, 1(4), 319-334.
- Scriven, M. (1994). Duties of the teacher. *Journal of Personnel Evaluation in Education*, 8(2), 151-184.

- Stemnock, S. K. (1969). *Evaluating teacher performance*. Educational Research Service Circular No. 3. Washington, DC: NEA 1969.
- Wagoner, R. L., & O'Hanlon, J. P. (Winter, 1968). Teacher attitude toward evaluation. *Journal of teacher education*, 19, 471-475.
- Wise, A. E., & Darling-Hammond, L. (1985). Teacher evaluation and professionalism. *Educational Leadership*, 42(4), 28-33.
- Wise, A. E., Darling-Hammond, L., McLaughlin, M., & Bernstein, H. T. (1984). *Teacher evaluation: A study of effective practices*. Santa Monica, CA: Rand.
- Wolf, R. L. (1971). *The role of the teacher in classroom evaluation*. M. A. thesis, University of Illinois.
- Zelanak, M. J., & Snider, B. C. (May, 1974). Teacher perceptions of the teacher evaluation process. *California Journal of Educational Research*, 110-134.

2

STANDARDS AND CRITERIA FOR TEACHER EVALUATION

Preamble: The Place and Importance of Standards

If the evaluation field is to achieve its potential contribution to any area, it must be capable of assessing all aspects of a discipline or system. This aspect includes the system's mission, guiding concepts, policies, constituents' needs, goals, plans, procedures, schedules, budgets, research design (including measuring devices and findings), communication, and personnel. Evaluation is a pervasive, and basically essential process for all aspects of society and its institutions as they strive for excellence, equity, and practicality in serving citizens. It follows that professional standards must be developed and used regularly to help ensure that evaluations attain the highest levels of quality and fairness in all their aspects. In education such standards must be employed to enhance and assess both systems used to evaluate teachers and evaluations of individual teachers.

The relevant history shows an accelerating increase in teacher evaluation, particularly during the last decade or so, and an associated need to improve teacher evaluation theory and procedures. This need has accompanied the growing realization that in order to educate students effectively, and to achieve related educational goals, educational institutions must use sound evaluation to select, retain, dismiss, professionally develop, and reward qualified personnel, particularly teachers, to achieve needed and planned outcomes. Further dimensions of this need include assessing the performance of teachers for many key purposes, such as making fair and sustainable decisions about promotion and tenure, and in other ways recognizing and rewarding merit; perceiving and remediating teaching or

teacher weaknesses; and developing and instituting an equitable, valid, and defensible case for terminating those who harm students and their learning. In addition, formative evaluation, one important role of teacher assessment, opens lines of communication between teachers and supervisors through the feedback provided from performance reviews. Moreover, it is incumbent on teachers to advocate and support sound teacher evaluation practice in the move to establish teaching as a respected field of professional practice. All these and other considerations in teacher evaluation indicate that it is both pervasive and important.

However, if teacher evaluation is to be convincing and fair, it must be underpinned by standards that help in detecting and correcting deficiencies in existing teacher evaluation systems and that offer educators, administrators, and board members widely shared principles for reviewing extant approaches, for developing and assessing new or improved approaches, for guiding these approaches to work beneficially, and for defending sound approaches against legal and other challenges. At all stages, teacher evaluation practices must be acceptable and credible. It is the function of professionally-developed standards to help achieve these ends.

However, despite the focus that has been placed on teacher evaluation by state mandates, and the realization by some school districts of its centrality, educational institutions too often have been ineffective in planning and implementing their personnel evaluation responsibilities. This failure was widely demonstrated during the late 1970s and increasingly in the 1980s. Nationally received publications such as those of the National Commission of Excellence in Education, *A Nation at Risk* (1983), and the *Carnegie Task Force on Teaching as a Profession* (1986) emphasized not only deficiencies in personnel evaluation in education, but also the damage such deficiencies caused. Numerous other instances can be cited indicating how widespread dissatisfactions with the quality of personnel evaluation was in education. Today there remains a dismaying disillusion by community groups and others vitally concerned about education who see educational evaluation as superficial and ineffective, particularly as it applies to teachers.

One poignant example comes from a research director of a large urban school district. He refuses to accept the findings of the district's teacher evaluation system, which indicates that 95 percent of the teachers are superior. This conclusion does not square with the fact that the achievement levels of most of the students in many of those teachers' classes, year after year, have remained embarrassingly substandard.

Personnel evaluation is not an easy business. In fact, development of the first set of standards for evaluation of personnel in education was not undertaken until the middle 1980s, about 4 years after the completion of the standards for program evaluation. The widely representative Joint Committee had explicitly excluded consideration of personnel evaluation when in 1975 it undertook the development of program evaluation standards, chaired by Daniel L. Stufflebeam. This avoidance arose from the realization that personnel evaluation is a highly controversial

activity, and that the Committee might doom even its attempt to develop program evaluation standards to failure if it did not postpone consideration of personnel evaluation. There was also the realization that extant personnel evaluation practices, though well entrenched, were replete with weaknesses such as failing to screen out unsuitable persons, provide evidence to withstand critical scrutiny, provide direction for staff professional development, and differentiate among staff to reward outstanding service. It was clear that if and when educational personnel evaluation standards were developed, a major, concentrated effort would be required.

Chapter 2.1: Development of Personnel Evaluation Standards by the Joint Committee

By the early 1980s, there was a clear need to address the mounting problems surrounding the evaluation of school personnel, and teachers in particular. State education policy makers seemingly concluded that the stalemate between and among school boards, teachers unions, and researchers that had impeded any progress to reform teacher evaluation would persist without direct intervention. A number of states quite suddenly mandated strong, "get tough" teacher evaluation systems. Typically, these were implemented prematurely and only added to the confusion, controversy, and litigation. The major undertaking by the Joint Committee during the 1980s (again chaired by Stufflebeam), culminating in the publication of *The Personnel Evaluation Standards* in 1988 was therefore most timely for education, and indeed, also for other professional institutions. These standards provided the education field with a set of guiding principles, the rigorous application of which should strengthen and add credibility to systems and practices of personnel evaluation as well as mitigate evaluation-related conflicts among the different interest groups.

Chapter 2.1 discusses the development of the Joint Committee's *Personnel Evaluation Standards*, and delineates areas where they focus, such as performance reviews, decision-making related to tenure and promotion, and staff development. The functions of personnel evaluation are encompassed in the Joint Committee's definition: *personnel evaluation is the systematic assessment of a person's performance and/or qualifications in relation to a professional role and some specified and defensible institutional purpose* (1988, p. 7). It had earlier defined a standard as *a principle commonly agreed to by people engaged in the professional practice of evaluation for the measurement of the value or the quality of an evaluation* (1981, p. 12).

Chapter 2.1 goes on to give a brief introductory statement about each of the four basic principles of sound evaluation—propriety, utility, feasibility and accu-

racy—and then makes a summary statement of the various standards conforming to each of these four general attributes. Suggestions are then offered for ways and means of applying the *Standards* to teacher evaluation systems. The chapter concludes with a number of key points concerning the applicability of the *Standards* for assessing and approving systems used to assess teachers.

Chapter 2.2: Criteria for Performance-Based Teacher Assessments: Validity, Standards, and Issues

In Chapter 2.2, Carol Anne Dwyer writes from her recent experience in leading the development of Praxis (the Educational Testing Service successor to NTE). She is concerned about the historic lack of defensible criteria, i.e. standards for assessing teacher competence, a matter already raised in PART 2. Praxis, like NTE, is the main evaluation device used by the states in teacher certification and thus must be grounded in a defensible definition of good teaching, which should be both appropriate for assessment purposes and also faithful to teaching as it is experienced by knowledgeable practitioners.

Dwyer explains that in teacher assessment, establishing assessment criteria is not synonymous with demonstrating content validity and that, instead, the design, development, and use of assessment criteria involve many aspects of validity related to the complexities of performance assessment. She observes that, among other important subject-matter and foundational knowledge and skills, Praxis is keyed to three different aspects of pedagogy: (1) content-specific pedagogical knowledge; (2) knowledge of general principles of teaching and learning that transcend different subject-matter disciplines; and (3) application of this knowledge and skill in actual classrooms.

Noting that these three aspects of pedagogy require different assessment methods, Dwyer then concentrates on the assessment of teaching practice in classrooms. In addressing the assessment of teaching, she notes that determining what to measure requires articulation of explicit standards that are defensible from educational, psychological, and measurement perspectives, and that take account of a particular view of teaching and learning. She recounts how the Praxis project brought home the reality that there remains a gap between the guidance provided by professional standards for performance assessments and the range of complex issues that have to be resolved in developing such assessments: for example, how to take account of the consequences of assessments in particular settings and how to consider a school district's particular philosophical and theoretical orientation to teaching and learning.

Dwyer usefully summarizes some of the recent progress in bridging the gap between professional standards and practice needs in performance assessment and

describes how the Praxis project both built upon and advanced the state of the art of standards for assessing teacher competence. Particularly, she analyzes and describes what ETS did to assess the following issues: creating a methodology for defining teacher assessment criteria; articulating a guiding conception of teaching and learning; examining and resolving diverse perspectives on teaching; balancing theory and practice in defining assessment content; finding the "right size" for criteria; and incorporating professional judgment in the assessment process.

Dwyer's chapter is on the cutting edge in both defining and addressing issues in the validation of teacher assessments. Her analysis also underscores both the complexity and huge expense involved in defining and applying teaching standards that (1) defensibly assess teaching in a variety of contexts, (2) continuously raise the standard of assessment development practice, and (3) lead to improved teaching.

References

- Carnegie Task Force on Teaching as a Profession. (1986). *A nation prepared: Teachers for the 21st century*. Hyattsville, MD: Carnegie Forum on Education and the Economy.
- Joint Committee on Standards for Educational Evaluation. (1981). *Standards for evaluation of educational programs, projects, and materials*. New York: McGraw-Hill.
- Joint Committee on Standards for Educational Evaluation. (1988). *The personnel evaluation standards*. Newbury Park, CA: Sage.
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative of educational reform*. Washington, DC: Government Printing Office, No. 065-000-00177-2.
- Scriven, M. (1988). Duty-based teacher evaluation. *Journal of Personnel Evaluation in Education*, 1(4), pp. 319-334.

Professional Standards for Assessing and Improving Teacher Evaluation Systems

As documented and discussed in Part 1, the history of teacher evaluation is characterized by a pervasive, unrelenting insistence throughout the U.S. society that state education agencies and school districts effectively evaluate teacher qualifications and performance; by many local- and state-level trial-and-error evaluation efforts to improve teacher evaluation; by many more ritualistic, perfunc-

tory, and, thus, benign evaluation programs; by widespread dissatisfaction with the quality and effectiveness of teacher evaluation practices; and by numerous invalidated models, forms, and state systems put forward to improve teacher evaluations. Clearly, there is a long-standing and important need to improve both the theory and practice of teacher evaluation.

As one means of confronting this need and advancing the practice of educational evaluation, a joint committee sponsored by the (then) 13 professional education societies—representing about 3,000,000 educators—developed professional standards for planning, operating, assessing, and validating educational personnel evaluation systems (Joint Committee, 1988). [Prior to developing *The Personnel Evaluation Standards*, the Joint Committee had developed the *Standards for Evaluations of Educational Programs, Projects, and Materials* (Joint Committee, 1981)]. Moreover, this Joint Committee on Standards for Educational Evaluation became a standing committee, set in place a mechanism and process for periodically reviewing and updating the Standards, and earned accreditation by the American National Standards Institute as the only body in the U.S. duly authorized to set standards for the practice of educational evaluation.

Therefore, the education field now has a set of principles for strengthening educational personnel evaluation practices and subjecting them to rigorous examination. Moreover, educators can have confidence that the Standards will be scrutinized on a regular basis and updated to keep pace with advancements in the technology of personnel evaluation.

We advocate very strongly that school board members, state education officials, educational administrators, teachers, evaluators, and others study *The Personnel Evaluation Standards* and develop the habit of systematically applying them in planning, implementing, assessing, and improving evaluation systems. Teacher evaluation, like any other area of professional service, must adhere to appropriate standards of good, acceptable practice. Now that education has produced *The Personnel Evaluation Standards*, educational policymakers and practitioners should use them to upgrade their teacher evaluation systems.

By summarizing and discussing *The Personnel Evaluation Standards* in this chapter, we aim to provide previously uninitiated readers with a basic introduction to the Standards and to provide all users of this book with a basis for examining the discussions of alternative evaluation approaches presented in subsequent chapters. Also, we recommend that readers who have not already done so study the full text of *The Personnel Evaluation Standards* and start using them on a regular basis.

Table 2-1. Types of Evaluations and Decisions Involved in Preparation, Licensing, Employment, and Professionalization of Teachers

Activities in Each Career Stage	Stages in the Career of a Teacher							
	Preparation		Licensing		Practice		Professionalization	
Entry	Evaluations of supply & demand	Decisions	Evaluations	Decisions	Evaluations	Decisions	Evaluations	Decisions
	Evaluations of recruitment programs	Ranking & funding training programs	Review of credentials	Approval to enter the certification process	Evaluation of staffing needs	Job definitions, job search	Examination of staff needs and institutional needs	Continuing education opportunities
	Assessment of applicants	Redesign of the programs			Evaluation of recruitment program	Program redesign	Assessment of needs & achievements of teachers	Approval of study leaves & special grants
		Selection of students			Evaluation of applicants	Selection of staff members	Assessment of basic qualifications for national certification	Participation in a national certification program
Participation	Intake evaluations	Planning student programs	Induction evaluation during a probationary year	Provisional state license	Comparison of job requirements & teacher competencies	Assignment	Intake evaluations	Designing individual education programs
	Evaluations of students' mastery of course requirements	Grades	Licensing	Partial qualifications for a license	Performance review	End of probation	Examination of competence	National certification
	Cumulative progress reviews	Counseling Remediation Counseling Revising student programs Termination				Tenure Merit pay Staff development Honors		
Exit	Final evaluation of students' fulfillment of graduation requirements	Graduation	Review of success in teaching for a designated period	Permanent or long-term license	Investigation of charges	Reduction in force	Participant achievement in continuing education	Qualifications for future leaves
	Exit interviews	Program review & improvement			Comparison of resources, staff needs, & staff seniority	Termination or sanctions	Examination of competence & aptitude	New assignments
	Follow-up survey	Program review & improvement			Performance review			

Personnel Evaluations for Which the Standards were Developed

The domain of application for *The Personnel Evaluation Standards* is depicted in Table 1. The table portrays personnel evaluation as an integral part of efforts by higher education, government, school districts, schools, and professional associations to prepare, license, engage, develop, and reward teachers and other educational personnel. The four main columns of the table relate to four career stages: *Preparation* (e.g., teacher education), *Licensing* (often called state certification), *Practice* (e.g., school teaching), and *Professional Development/Advancement* (e.g., study leaves, inservice education, national certification). The three rows of the table divide each career stage into *Entry* activities (e.g., selection of a candidate for entry into one of the four stages depicted on the main horizontal dimension, such as a teacher preparation program), *Participation* (e.g., preparation to become a teacher or actual classroom teaching in a school), and *Exit* (e.g., graduation from a teacher education program or retirement or termination from a classroom teaching position). The four main columns of the table are subdivided into *Decisions* and *Evaluations* that are involved in the Entry, Participation, and Exit activities of each of the 4 career stages of a teacher.

The matrix is designed to encompass all the decisions and associated evaluations involved from the beginning of a teacher's preparation and extending throughout the teacher's career. As such, it helps to identify the range of evaluation criteria and methods needed to meet all requirements of teacher evaluation.

The full range of evaluations depicted in the matrix are important to staffing schools successfully with qualified and effective teachers. It cannot be overemphasized that the totality of teacher evaluation activity requires careful attention. Poor decisions in selecting, certifying, and tenuring teachers can have long term, negative consequences for students, schools, and teachers themselves. Education authorities need to ground all their decisions about prospective and practicing teachers in sound evaluations. The evaluation criteria, standards, and methods used must be valid and systematically applied. Only then can evaluation realize its full potential to impact the quality of teaching and to help assure that competent teachers effectively assist all students to learn and develop their capacities.

Also, the teaching profession needs to support and exercise systematic teacher evaluation as a means of improving professional accountability. The present public image of teaching is poor. This is due in part to a wide perception that education authorities have employed weak standards in preparing, placing, and retaining teachers and have not subjected classroom teaching to rigorous, consequential evaluation. By increasing the systematic employment of sound evaluation and rigorous decision standards at every career stage of teaching, the education establishment undoubtedly would improve the quality of the teaching force. This in turn should improve the professional credibility of teaching and make the public more

willing to support higher teaching salaries. By regularly conducting sound teacher evaluation, the institutions responsible for the different stages of teacher development and deployment would be able to provide the public with pertinent evidence to answer questions about the quality of teacher selection, preparation, and performance.

In sum, paying attention to all the evaluations and decisions referenced in the matrix should help educators address a range of important concerns:

- assuring that the strongest possible efforts are made to select promising teacher education students
- systematically preparing student teachers for teaching service
- thoroughly assessing student teachers' fulfillment of preservice education requirements before licensing them for long service
- carefully examining the qualifications of teachers for particular assignments
- monitoring the progress of teachers and providing feedback for improvement and professional growth
- making appropriate decisions concerning teacher retention, promotion, tenure, and recognition

The cell entries in the matrix show that the wide range of decisions and associated evaluations in teacher evaluations are of 3 types. A few are **program evaluations**, e.g., evaluations of recruitment programs. Others are **student evaluations**, such as evaluation of teacher education students' mastery of university courses. Finally, most of the evaluations identified in the matrix fit the common view of **personnel evaluation**, i.e., assessing the qualifications and performance of individual educators as a basis for licensing, appointment to a position, supervision, staff development, promotion, tenure, merit pay, and national certification. All evaluations considered crucial for developing and engaging effective teachers should adhere to appropriate professional standards.

The Personnel Evaluation Standards are focused on the educational personnel evaluations identified in the matrix. As shown, given categories of evaluation often provide information for use in making different decisions. This is especially true for performance reviews, the main topic of this book, that provide information for decisions about tenure, promotion, merit pay, and other awards, as well as counseling for staff development. Because given evaluations typically have several potential uses, all parties to an evaluation must enter into it with a clear idea of who will receive the information and how it will be used: it is unfair to decide after the fact that an evaluation intended, for example, for private feedback and staff development, will also be used to determine one's tenure or termination.

The Personnel Evaluation Standards have broad application to the full range of teacher evaluations identified in Table 2-1. School districts share a need for sound evaluations by which to choose new teachers, to examine teaching performance, to counsel teachers, to recognize extraordinary performance, to make tenure and

promotion decisions, and, when needed—in order to stay within budget or protect the interests of students—to make reduction-in-force or other termination decisions. *The Personnel Evaluation Standards* are presented as a single set of principles that educators and their constituents can use to plan or evaluate evaluation systems for the full range of personnel actions shown in Table 2-1.

The Joint Committee intends, further, that the Standards be applied to all professional roles in schools, post secondary institutions, and other institutions that have a primary responsibility to educate. They also intend that the Standards will be applicable to a broad range of techniques, including observation, interview, applied performance tests, licensing tests, simulations, professional skills tests, assessment centers, portfolio development, supervisor assessment, peer assessment, self assessment, and student assessment.

Guiding Definitions

According to the Joint Committee, personnel evaluation is the systematic assessment of either or both a person's performance and qualifications in relation to a role and some specified and defensible institutional purpose. The Committee defined a standard as "a principle commonly agreed to by people engaged in the professional practice of evaluation for the measurement of the value or the quality of an evaluation."

The Committee presented *The Personnel Evaluation Standards* at the level of elaborated general principles, with a wide range of illustrations constructed to help users see how to apply the Standards to the various types of personnel evaluations identified in Table 2-1. The Committee emphasized that general principles are adequate for providing direction for improvement, and that they avoid oversimplifications and leave room for creative and locally responsive evaluation procedures. However, in order to assure that the Standards will have practical utility, the Committee provided concrete suggestions and examples concerning both how to meet each standard and what common errors should be avoided.

The Four Basic Principles of Sound Evaluation

Sustaining the position it adopted in its program evaluation standards, the Joint Committee grounded its development of *The Personnel Evaluation Standards* in a fundamental proposition. It is that all evaluations should have four basic attributes: propriety, utility, feasibility, and accuracy.

Propriety Standards To satisfy the condition of propriety, teacher evaluations should be ethical and fair to all parties—to the students and their parents as well as to the teachers, administrators, and evaluators. The Propriety Standards reflect the fact that personnel evaluations often violate or fail to address certain important ethical and legal principles. The primary principle is that schools exist to serve students; therefore, personnel evaluations should concentrate on determining whether educators are effectively meeting the needs of students. Moreover, the evaluations should provide direction for helping evaluatees to improve service and, when necessary, should provide a basis for terminating the appointments of educators who prove to be persistently incompetent and/or unproductive. Overall, the Propriety Standards require that evaluations be conducted legally, ethically, and with due regard for the welfare of students, other clients, and educators.

The summary statements for the five Propriety Standards are as follows:

P-1 Service Orientation

Evaluations of educators should promote sound education principles, fulfillment of institutional missions, and effective performance of job responsibilities, so that the educational needs of students, community, and society are met.

P-2 Formal Evaluation Guidelines

Guidelines for personnel evaluations should be recorded and provided to employees in statements of policy, negotiated agreements, and/or personnel evaluation manuals, so that evaluations are consistent, equitable, and in accordance with pertinent laws and ethical codes.

P-3 Conflict of Interest

Conflicts of interest should be identified and dealt with openly and honestly, so that they do not compromise the evaluation process and results.

P-4 Access to Personnel Evaluation Reports

Access to reports of personnel evaluation should be limited to individuals with a legitimate need to review and use the reports, so that appropriate use of the information is assured.

P-5 Interactions with Evaluatees

The evaluation should address evaluatees in a professional, considerate, and courteous manner, so that their self-esteem, motivation,

professional reputations, performance, and attitude toward personnel evaluation are enhanced or, at least, not needlessly damaged.

Utility Standards. The Utility Standards are intended to guide evaluations so that they will be informative, timely, and influential; in other words, so that they are not simply annual perfunctory, ritualistic exercises of no importance, as so often has been the case in U.S. schools. The key point of these standards is to assure that evaluations provide information of use to individuals and groups of educators in examining and improving their performance. The Utility Standards also require that evaluations be focused on predetermined uses, such as informing selection and promotion decisions or providing direction for staff development; be addressed to and accessed by predesignated users; and be conducted by persons with appropriate expertise, credibility, objectivity, and authorization. In general, these standards view personnel evaluation as an integral part of an institution's ongoing effort to select outstanding staff members through timely and relevant evaluative feedback; to encourage and guide them to deliver high quality service; and by identifying and helping to terminate unproductive staff members, to effectively safeguard the welfare and educational interests of students.

The Utility Standards should be especially welcome to teachers, supervisors, and directors of inservice training who see their district's performance review system as only ritualistic and not helpful, or, worse, demoralizing and counterproductive. By applying the Utility Standards, a school district or other educational institution would be guided to clarify intended uses and associated information requirements of its evaluation system and to implement appropriate steps to ensure that the system addresses relevant questions, communicates useful reports, and provides direction for improvement. The main point of the Utility Standards is to insure that evaluations contribute constructively to helping educators deliver excellent service.

The summary statements for the five Utility Standards are as follows:

U-1 Constructive Orientation

Evaluations should be constructive, so that they help institutions to develop human resources and encourage and assist those evaluated to provide excellent service.

U-2 Defined Uses

The users and the intended uses of a personnel evaluation should be identified, so that the evaluation can address appropriate questions.

U-3 Evaluator Credibility

The evaluation system should be managed and executed by persons

with the necessary qualifications, skills, and authority, and evaluators should conduct themselves professionally, so that the evaluation reports are respected and used.

U-4 Functional Reporting

Reports should be clear, timely, accurate, and germane, so that they are of practical value to the evaluatee and other appropriate audiences.

U-5 Follow-Up and Impact

Evaluations should be followed up, so that users and evaluatees are aided to understand the results and take appropriate actions.

Feasibility Standards. The Feasibility Standards are grounded in the fact that personnel evaluations are performed in institutions that have limited resources and are influenced by a vast range of dynamic national, state, community, and institutional forces. Accordingly, the Feasibility Standards require evaluation systems that are easy to use, adequately funded, frugally managed, and politically viable.

The summary statements of the three Feasibility Standards are as follows:

F-1 Practical Procedures

Personnel evaluation procedures should be planned and conducted, so that they produce needed information while minimizing disruption and cost.

F-2 Political Viability

The personnel evaluation system should be developed and monitored collaboratively, so that all concerned parties are constructively involved in making the system work.

F-3 Fiscal Viability

Adequate time and resources should be provided for personnel evaluation activities, so that evaluation plans can be effectively and efficiently implemented.

Accuracy Standards. The Accuracy Standards call for evaluations that are based on dependable information about relevant qualifications or performance of a teacher or other educator. These standards require that the obtained information be job related, technically defensible, and appropriately interpreted. The overall rating of a personnel evaluation against the Accuracy Standards gives a good measure of the evaluation's validity.

These standards particularly demand that evaluations be grounded in the duties of the teacher or other educator. Accordingly, these standards call for variables to be derived from a valid description of the person's job. Simply showing that a personal characteristic—such as teaching style, quantitative aptitude, personal appearance, age, sex, or race—is correlated with student achievement does not justify using the characteristic to measure and judge either the qualifications or performance of a teacher or other educator. As Scriven (1988) has argued, to do so not only risks prejudicial treatment of individuals, but, since the correlations are based on group data and are never perfect, such practice also produces invalid assessments of persons who rate low or "adversely" on the variable but do well on the job, or vice versa. The Joint Committee's field tests clearly indicated that many personnel evaluation systems need to be improved in how well they define jobs, how carefully they consider environmental influences, how validly they measure job qualifications and performance, and how effectively they control for various kinds of bias.

The summary statements for the eight Accuracy Standards are as follows:

A-1 Defined Role

The role, responsibilities, performance objectives, and needed qualifications of the evaluatee should be clearly defined, so that the evaluator can determine valid assessment criteria.

A-2 Work Environment

The context in which the evaluatee works should be identified, described, and recorded, so that environmental influences and constraints on performance can be considered in the evaluation.

A-3 Documentation of Procedures

The evaluation procedures actually followed should be documented, so that the evaluatees and other users can assess the actual, in relation to intended, procedures.

A-4 Valid Measurement

The measurement procedures should be chosen or developed and implemented on the basis of the described role and the intended use, so that the inferences concerning the evaluatee are valid and accurate.

A-5 Reliable Measurement

Measurement procedures should be chosen or developed to assure reliability, so that the information obtained will provide consistent indications of the performance of the evaluatee.

A-6 Systematic Data Control

The information used in the evaluation should be kept secure, and should be carefully processed and maintained, so as to ensure that the data maintained and analyzed are the same as the data collected.

A-7 Bias Control

The evaluation process should provide safeguards against bias, so that the evaluatee's qualifications or performance are assessed fairly.

A-8 Monitoring Evaluation Systems

The personnel evaluation system should be reviewed periodically and systematically, so that appropriate revisions can be made.

Applying the Standards to Teacher Evaluation Systems

The Personnel Evaluation Standards are a systematically developed and widely endorsed basis for assessing and improving systems for evaluating the qualifications and performance of teachers and other educators. It is reasonable to expect school districts, state education departments, universities, and other educational agencies to use the Standards as a checklist of basic requirements and associated procedural suggestions, both to assure that their personnel evaluation systems are sound and to make needed or desirable improvements.

Steps of use in applying the Standards were recommended by Stufflebeam and Brethower (1987) and adapted by the Joint Committee for inclusion in *The Personnel Evaluation Standards* book. An updated version of these steps follows:

1. Study the Standards

- Consider adopting the Standards as the basic reference by which to examine and promote quality in personnel evaluation.
- Make copies of the Standards available for study and use by school board members, administrators, teachers, and other interested parties.
- Conduct workshops aimed at teaching the Standards and illustrating their use.
- Appoint a Study Group on applying *The Personnel Evaluation Standards*. Ideally, it should reflect the perspectives of the Joint Committee on Standards for Educational Evaluation (i.e., teachers, evaluators, administrators, policy board members, curriculum developers, personnel specialists, psychologists, research and testing personnel, counselors, and specialists in educational law).
- Charge the Study Group to apply the Standards to the personnel evaluation system, to report which standards are met and not met, to identify

deficiencies to be overcome, and to issue recommendations for strengthening or replacing the system.

2. Clarify the purposes of the evaluation system by determining
 - Whose work is to be evaluated?
 - Why should the evaluations be done?
 - Who should and will use the findings?
 - What decisions will be determined or affected and/or what types of actions are evaluatees and managers expected to take in response to evaluation reports?
 - Should the evaluation(s) focus on qualification, performance, and/or effectiveness?
 - What impact is the evaluation system intended to have?
3. Describe the personnel evaluation system to be examined
 - Assemble relevant documents (e.g., personnel policies, negotiated agreements, job descriptions, letters of appointment, rating and reporting forms).
 - Describe or outline the evaluation system (delineating how the evaluations are staffed, the qualifications of the evaluators, the training and orientation they receive, the pertinent policies, the evaluation purposes and questions, the measurement variables and procedures, the procedures for organizing and keeping the data secure, the procedures for weighting and analyzing the findings, the reporting formats and schedule, the uses of findings, practices in providing follow-up support, the management system, provisions for appeal and review of findings, and the practices in evaluating and improving the evaluation system).
4. Apply the Standards
 - Engage the Study Group in applying the Standards in assessing the extent to which they believe each standard has been satisfied.
 - For each standard the study group should list strengths and weaknesses of the personnel evaluation system and make an overall decision whether the standard is met, partially met, or not met.
5. Decide what to do about the results
 - Engage the Study Group to discuss the results and develop recommendations for improving the system.
 - Share the results with the professional staff and board of the institution and obtain their input.
 - Develop a general plan for the institution to use in improving the evaluation system.

The preceding is an outline of five general steps to follow when using the Standards to examine and upgrade a personnel evaluation system. The process is

functional and the Standards applicable, as evidenced by the field tests of the draft Standards. Those tests showed that institutional committees can work through a process like the one described above within a three to six months period and as a result produce a shared plan for improving a personnel evaluation system. For an excellent example of such an application, see the article entitled, "Review of Personnel Evaluation Systems" (Reineke, Willeke, Walsh, & Sawin, 1988). In that article, the authors reported how they used the Standards to review and revise the teacher evaluation system employed by the Lincoln, Nebraska, Public Schools.

Applicability of the Standards

Experiences thus far recorded in applying the Standards, including about 50 field tests, lead to the following conclusions concerning their applicability for assessing and improving systems for evaluating teachers and other educators:

1. Unless the Standards are systematically applied, evaluation committees may commit many serious errors, e.g., allowing conflict of interest to influence and discredit results, wasting time and resources in gathering data that won't be used, engaging untrained evaluators, allowing political factors to distort process and findings, or producing a controversial or legally vulnerable evaluation system.
2. Collaborative and professional planning and implementation of the personnel evaluation system will help to ensure commitment to its credibility, propriety, and utility. Allow time to engage in reflective discussion with interested stakeholders. The Political Viability standard provides a useful perspective and set of suggestions for collaboratively developing evaluation systems.
3. To assure that an evaluation system will be accepted, effectively used, and legally viable, an institution must ground its evaluation work in clear and defensible institutional policies and guidelines. A clear understanding of, and commitment to, the use of the Standards by all concerned can foster clear and accepted evaluation guidelines, collaborative development of evaluation procedures, predetermined uses of given evaluations, controlled and appropriate access to personnel evaluation files, clear and timely reporting, use of results to counsel and assist evaluates, search for and control of bias, defense against legal attack, and periodic review and improvement of the evaluation system. The Formal Evaluation Guidelines standard is focused particularly on the issues and appropriate steps to developing sound institutional policies and guidelines on personnel evaluation.

4. The Service Orientation and Constructive Orientation standards are crucial counter balancing values that must be invoked consistently if evaluations are to assure that students receive competent service and that individual educators are continually guided to improve their skills and services.
5. The Defined Role standard is fundamental to operating a sound personnel evaluation system. The validity of criteria and data used in any personnel evaluation must reference a valid definition of the evaluatee's role.
6. Before determining that each educator must be evaluated each year, institutional authorities should carefully determine whether and how such annual evaluations would be used. Evaluations should be done only when there is a clear provision for using the results; to do evaluations whose results predictably will not be used is both wasteful and demoralizing. Careful attention to the Defined Uses standard will help institutions to avoid problems of this nature.
7. Conflict of interest is an ever-present issue in evaluating teacher performance. Practically, the evaluator(s) will almost always be the principal, a group of peers, the students, or some combination. Therefore, the evaluatee and evaluator work in close proximity and to some degree must collaborate in achieving their mutual and shared objectives. In close working relationships, it is natural for friendships, animosities, and mutual dependencies to develop, and these can influence how one person evaluates another's performance. By paying careful attention to all of the standards, and particularly to the Conflict of Interest, Valid Measurement, and Bias Control standards, the institution can do much to overcome problems of conflict of interest.
8. A bad habit that is pervasive in the education field is selecting and using a standardized, simplistic form as the means of performance evaluation, even though the form has gone through no validation process. Such forms are usually chosen because they have "face validity" and are easy to use. Typically, their use is neither respected nor productive and certainly could not be defended when matched against the requirements of *The Personnel Evaluation Standards*. Careful attention to the Defined Role, Work Environment, Valid Measurement, and Practical Procedures standards will aid the education field to move productively away from the use of invalid evaluation forms and procedures.
9. Ability to assess performance of teachers and other educators accurately and effectively requires a significant institutional investment of necessary resources and time, effective training of evaluators, consistency and equity in application of the process, constructive use of findings, and periodic review and improvement of the evaluation system. Institutions must invest adequately in personnel evaluation systems if those systems are to produce valid and reliable information and to have a positive impact on performance.

These investment issues are addressed especially by the Evaluator Credibility, Fiscal Viability, Systematic Data Control, and Monitoring Evaluation Systems standards.

10. Personnel evaluations often lead to legal proceedings over the issue of basic due process requirements. Many of the requirements of basic due process will be met if the Standards are followed. The Functional Reporting standard should be of assistance in meeting the notice requirement of due process. The Defined Role standard enhances the defensibility of a determination that an employee failed to fulfill her or his responsibilities. Also, attending to the Practical Procedures, Valid Measurement, Reliable Measurement, Systematic Data Control, and Bias Control standards helps to meet some of the procedural requirements of due process.

Conclusion

This chapter has described the response by the education profession, through the work of the Joint Committee on Standards for Educational Evaluation, to a critical need in education, the need for professionally defined and widely endorsed standards for use in examining and improving personnel evaluation systems, including especially the systems used to assess the qualifications and performance of teachers. Personnel evaluation systems must be subjected to periodic review and development. Policies, contracts, laws, evaluation technology, job assignments, key actors, teacher morale, and other conditions can change and consequently may require changes in the evaluation system. *The Personnel Evaluation Standards* provide a comprehensive checklist of principles, issues, and recommendations to consider when conducting such reviews. The Documentation of Procedures and Monitoring Evaluation Systems standards are particularly focused on continually assessing and improving evaluation systems.

To assure that personnel evaluation will bring about outstanding educational services, the education profession and states must assess the qualifications and performance of educators at all stages of their careers. Current programs to assess both applicants and graduates of teacher education programs, to replace the National Teachers Examination with the ETS Praxis series, and to develop a rigorous program of national certification for outstanding teachers are consistent with this position. These and other related evaluation programs should pay special heed to the Service Orientation and Constructive Orientation standards.

References

- Joint Committee on Standards for Educational Evaluation. (1981). *Standards for evaluations of educational programs, projects, and materials*. New York: McGraw-Hill.
- Joint Committee on Standards for Educational Evaluation. (1988). *The personnel evaluation standards*. Newbury Park, CA: Sage.
- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards*, 2nd edition. Thousand Oaks, CA: Sage.
- Reineke, R. A., Willeke, M. J., Walsh, L. H., & Sawin, C. R. (1988). Review of personnel evaluation systems: A local application of the *Standards*. *Journal of Personnel Evaluation in Education*, 1, 373-378.
- Stufflebeam, D. L., & Brethower, D. M. (1987). Improving personnel evaluations through professional standards. *Journal of Personnel Evaluation in Education*, 1, 125-155.

Criteria for Performance-Based Teacher Assessments: Validity, Standards, and Issues

By Carol Anne Dwyer

Introduction

Identifying appropriate content for teacher performance assessment criteria is a complex conceptual and empirical task and one that has close connections to validation theory and to both traditional and emerging testing standards. Perhaps the major unresolved validity issue from both the testing specialist's and the lay person's point of view is the absence of technically, logically, educationally, and ethically defensible criteria for good teaching. The lack of such criteria has been the focus of sharp criticism of teacher assessment for many years and has remained a central issue in establishing the validity of any teacher assessments despite the difficulties it presents. This chapter will take the point of view that establishing assessment criteria is not synonymous with demonstrating content validity. Rather, the design, development, and use of assessment criteria involve many aspects of validity and are related to many testing standards that now exist or that are currently being proposed for performance assessments. This chapter discusses validity and standards-related issues raised in developing teacher performance assessment

criteria, drawing on a six-year research and development effort to create assessments for licensing beginning teachers, *The Praxis Series: Professional Assessments for Beginning Teachers*¹. This effort culminated in the creation of a set of classroom performance assessments.

A central concern in creating these assessments was one that has plagued teacher assessment for decades and that has often been described as intractable: how to define good teaching in a way that is appropriate for assessment purposes and yet remains faithful to teaching as it is experienced by knowledgeable practitioners. In the course of this project, a methodology was created to address this problem; a conception of teaching was articulated and operationalized for assessment purposes through a national effort to create a set of assessment criteria; and, in the process, issues related to standards and validity were identified and resolved for purposes of design, development, and use. These policy and values issues have implications for teacher educators, teaching practice, and research, as well as for assessment development and use in teaching and other performance assessment contexts.

Note on Assumptions and Definitions. This chapter focuses on performance assessment, and many of the examples are taken from the development of *The Praxis Series: Professional Assessments for Beginning Teachers*TM, which has been described in detail elsewhere (Dwyer, 1992; 1993a; 1993b). It is important to note, however, both that the issues discussed in this paper are not limited in their applicability to a particular set of assessments, and that performance assessment of classroom teaching does not exist in isolation in *The Praxis Series*. In the development process for *The Praxis Series*, thorough examination of the assessment needs for licensing beginning teachers identified, among other important subject-matter and foundational knowledge and skills, the need for assessment of three different aspects of pedagogy: (1) content-specific pedagogical knowledge, (2) knowledge of general principles of teaching and learning that cut across a variety of subject matter disciplines, and (3) application of this knowledge and skill in the context of the actual classroom. In addition to differences in content, these three aspects of pedagogy require different methods of assessment. The focus of this chapter will be on the last of these three aspects of pedagogy, the assessment of actual teaching practice, in its natural classroom context. It should be noted, however, that this focus does not imply any order of pedagogical or methodological merit. All three are important aspects of teaching; assessment of one type of pedagogy does not in

¹ Copyright ©1993 by Educational Testing Service. All rights reserved. EDUCATIONAL TESTING SERVICE, ETS, and the ETS logo are registered trademarks of Educational Testing Service. THE PRAXIS SERIES: PROFESSIONAL ASSESSMENTS FOR BEGINNING TEACHERS and its design logo are trademarks of Educational Testing Service.

itself allow inferences about the other types; and no assessment methodology is intrinsically superior to other methodologies for every purpose.

Validity and Standards for Performance Assessments The adequacy of efforts to define what should be measured about teaching cannot be meaningfully determined without articulation of explicit standards in a comprehensive frame of reference that encompasses issues traditionally of concern to education, psychology, and measurement. As will be argued below, this evaluation must also be contextualized with respect to a particular view of teaching and learning. Modern validity theory (Cole & Moss, 1989; Messick, 1989, 1992; Moss, 1992; Tittle, 1989), in its emphasis on a broad context for establishing assessments' validity, provides a great deal of conceptual guidance in considering the validity of assessments of complex performances such as teaching and on technical aspects of determining their content. Such work on validation theory highlights the measurement implications of the interconnections that exist within the whole system of which assessment is a part, and thus the value inherent in direct measurement of performance where this is feasible. This emphasis on context and consequences means looking at assessments to see, in broad terms, whether they do harm or good.

Although the theoretical basis for a broad, construct view of validity, including its emphasis on consequences of assessment, is now very widely accepted by the educational and psychological measurement communities (Moss, 1992), and is in fact codified in the most recent revision of the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 1985), there is still a considerable gap between the literature on validity and the very diverse and challenging set of issues faced in developing teacher performance assessments in a high-stakes environment. Researchers and developers in this area have lacked definitive guidance from the literature on such important issues as taking into account the classroom subject matter and human variables in teacher assessments (how to deal with contextual differences) and on the validity implications of creating assessments that take as their starting point the view that learning is an active process of constructing meaning from prior experiences (assessments with a constructivist foundation). These issues are critical to determining the criteria by which teaching performance will be assessed, but a leap is required to bridge from the solid theoretical base in the research literature to actual research and development practice.

Perhaps the most widely used and cited standards are those developed by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education, *Standards for Educational and Psychological Testing* (1985). These standards are very specific in putting forth a coherent view of validity in assessment development, but are at

a level of generality that still leaves a considerable gap between these theoretically-grounded standards and some important issues facing developers of teacher performance assessors. These standards are periodically updated and a new revision of them has just begun, so there is reason to hope that performance assessment issues will be addressed in more detail soon.

This is not to say that useful literature on standards, expectations, and aspirations in performance assessment and personnel evaluation is completely lacking—it has in fact begun to bridge the gap between validity theory and development practice. Moss (1992) addresses this issue as part of a comprehensive review and analysis of recent changes in validation theory and their implications for performance assessment. Although her analyses specifically concern the organization of validity inquiry, they shed light on ways in which various authors have begun to suggest, directly or indirectly, areas in which validation theory implies standards for performance assessment.

Miller and Legg (1993) also discuss both evidential and consequential aspects of validity as they relate to performance assessments and issues raised by various types of performance assessments, but they do not propose standards *per se*. A number of standards with at least arguable relevance to teacher performance assessments have been proposed, and many touch specifically on the issue of defining the target performance. Papers that concern issues relevant to standards for performance assessment are abundant, but few are intended to suggest specific standards. For example, Haertel (1991) discusses desirable characteristics of performance assessments and raises a set of technical issues that imply, but do not actually propose, evaluative standards. Similarly, Frederiksen and Collins (1989) in their paper describing a systems approach to testing emphasize the significance of context in evaluating assessments and suggest a number of principles for designing educational assessments that will have a positive effect on the system of which they are a part. Their analysis proceeds in ways that are suggestive of standards, and in fact briefly advocates the importance of developing four types of standards: directness of measurement; scope or inclusiveness of what is measured; reliability; and transparency, or meaningfulness of assessment criteria to test takers. They do not specifically limit the applicability of these standards, or their discussion in general, to performance assessment.

Work also exists in related areas, such as that of Claxton, Murrel, and Porter (1987) on assessments of college students' complex performances, that approaches proposing standards in the areas with relevance to standards for teacher assessment, such as curricular relevance, utility to decision makers, consistency with educational goals, feasibility, faculty involvement, improved student learning, and perceived value to students. This view of standards for student assessments maps well onto those proposed for performance assessments in general by Linn, Baker, and Dunbar (1991).

Linn, Baker, and Dunbar have proposed a set of expectations and validation criteria for performance assessments that seem directly applicable to high-stakes teacher performance assessments. Linn, Baker and Dunbar, unlike other authors noted above, are explicit in their intention to propose evaluative standards. Some of these proposed standards are easily linked to validity theory; others go beyond validity into areas of practical concern. The eight standards that they propose concern (1) consequences; (2) fairness; (3) transfer and generalizability; (4) cognitive complexity; (5) content quality; (6) content coverage; (7) meaningfulness; and (8) cost and efficiency.

Quellmalz (1991), perhaps best known for her work in writing, where performance assessments have a long history, proposes evaluating performance assessment criteria on the basis of six related criteria: significance, fidelity or appropriateness to the context, generalizability, developmental appropriateness, accessibility, and utility.

At a different level of specificity and primarily to focus their own research and development, the National Board for Professional Teaching Standards (NBPTS) has committed itself to a set of aspirational goals for assessing accomplished teachers that relate closely to those proposed by Linn et al. and that similarly go beyond areas generally considered within the purview of validity. NBPTS has committed itself to creating assessments for accomplished teachers that are (1) administratively feasible, (2) professionally credible, (3) publicly acceptable, (4) legally defensible, and (5) economically affordable (National Board for Professional Teaching Standards, 1991, p. 53).

Educational Testing Service (ETS) has developed *ETS Standards for Quality and Fairness* (1987), which specifically reference the AERA et al., *Standards* (1985). The *ETS standards* cover areas pertinent to performance assessments such as validity, test development, test administration, test use, and score interpretation. These *Standards* are currently being revised to reflect recent developments in alternative assessment. Additional project specific guidelines were also developed as part of The Praxis Series: Professional Assessments for Beginning Teachers™ (Educational Testing Service, 1992, 1993).

The Joint Committee on Standards for Educational Evaluation (a project of 14 educational, psychological, and measurement organizations, chaired by Daniel L. Stufflebeam of Western Michigan University), with the clear intention of providing operational guidelines for users and developers of personnel evaluations in education, created a set of standards that are specifically intended to apply to performance assessments of practicing teachers and other educators (1988). These standards are organized into groups of standards concerning propriety, utility, feasibility, and accuracy. Many practical issues are addressed in the first three categories; the fourth comes closest to addressing validity concerns. These standards are noteworthy in their explicit attention to contextual factors in assessments, although these are not

labelled as validity concerns in the Joint Committee's framework. For example, these standards require evaluation on the basis of a service orientation; that is, the extent to which the educational system as a whole benefits from the assessment activity. This standard is located in the section on propriety standards, however, rather than in the section that includes validity and bias control standards. The standards also address consideration of the work environment and its characteristics as part of the basis for reaching conclusions about an educator's effectiveness.

In sum, the body of literature on validity, standards, expectations, and aspirations, while highlighting important conceptual and practical issues, creates a heavy burden of interpretation and extrapolation for developers of performance assessments. There are clearly concerted exhortations in this literature to assessment developers and evaluators, albeit in varying voices, to take the broad consequences of assessment into account; to incorporate elements of context into the assessment process; to focus accurately on the knowledge and skills about which one wishes to draw inferences; and to include in the assessments the full range of content about which inferences are to be drawn. Despite this high level of agreement in principle, much discretion, and much responsibility, is necessarily left to individual creators of assessments and to those who evaluate their efforts.

In the development of large-scale teacher assessments described here, certain validity-related issues emerged that appear to have relevance to target performance definition in a wide range of performance assessment applications (as well as to a number of educational research and teacher education concerns), but that currently lack specific guidance from the standards, expectations, and aspirations literature. In the next section of this chapter, some of these issues and their resolution in defining good teaching in the development of the Praxis III: Classroom Performance Assessments will be discussed. The issues discussed are creating a methodology for defining teacher assessment criteria, articulating a guiding conception of teaching and learning, resolving diverse perspectives on teaching, balancing theory and practice in defining assessment content, finding the "right size" for the criteria, and understanding the role of professional judgment in the assessment process.

Articulating a Conception of Teaching and Learning

Assessment criteria judged to be technically, professionally, and legally defensible must proceed from an explicit conception of teaching and learning. It is critically important for the purposes of teacher performance assessment to formulate a guiding conception that explicitly recognizes the connection between teaching and learning. Not only is it impossible to discuss what is fundamental about one without considering the other, but joint consideration of the two facilitates evaluation of their linkages at the level of standards and assessment criteria. This means that an

important validation concept, impact of assessment on the total system of which it is a part, can be evaluated.

An effective guiding conception of teaching and learning should thus be explicit about where the assessment criteria come from and should lead directly to implications for assessment development and use. It should lead to inferences about both the content of the assessments and the methods used to collect data. For these reasons, it should also be articulated early in the development process, before final design decisions are made.

For example, the conception of teaching and learning that governed the development of the Praxis III: Classroom Performance Assessments (Dwyer & Villegas, 1993) includes premises such as that effective teaching requires both action and decision making and that learning is a process of the active construction of knowledge. This guiding conception also makes explicit the belief that because good teaching is dependent on the subject matter and the students, assessments should not attempt to dictate a teaching method or style that is to be applied in all contexts (see Scriven, 1990, for an elaboration of the rationale for this point). That is, effective teachers adapt instruction to the needs of the students and the situation rather than rigidly follow fixed scripts. The complexity of teaching thus requires making thoughtful decisions, then putting them into action. Because classroom life is complex and varies with regard to students and subject matter, among other things, teachers need to develop an instructional repertoire and skill in selecting from this repertoire procedures that are appropriate for the particular situation. Because this conception of teaching holds that good teaching decisions can be made only with reference to the subject matter and the students being taught, the cultural characteristics of the students (including students' ethnicity, gender, socioeconomic background, and exceptionalities) are extremely salient. This view of teaching is thus linked to a view of learning that is active and constructivist in that it holds that teachers build on the individual student's existing knowledge, which is in turn linked to the student's cultural resources.

This conception of teaching leads directly into design decisions, thus linking educational and psychological theory with measurement practice. For example, the emphasis on teacher decision making and on the importance of context in evaluating that decision making, strongly implies the value of data gathering in the actual classroom setting, as opposed to simulations. A second implication is that the assessment should include opportunities for the assessor and assessee to interact about the teaching event that is being considered. A third implication is that because there is no "one right answer" to the question of what is good teaching (because teaching is seen as inherently context sensitive), the scoring of the assessments must allow for multiple forms of acceptable "answers," while clearly articulating what constitutes unacceptable professional practice.

A fourth implication related to the complexity of teaching is that the assessment process, relying as it does on consideration of a complex set of data, will require substantial professional judgment to implement. Assessors should thus be experienced professionals who have been trained to reach a common understanding of the assessment criteria and other considerations for applying them. The judgment required of these assessors will be discussed in a later section of this paper.

Another aspect of this guiding conception of teaching is a set of value assumptions that are related both to proposed standards and aspirational statements for performance assessments. These values assumptions include the following:

- Teacher assessments should contribute to the equitable treatment of all teachers and their students.
- The assessment should be a learning experience for all of the participants, the assessor as well as the beginning teacher.
- Specific and professionally meaningful standards of teaching knowledge and practice can be developed and assessed.
- Assessments must be geared to the prospective teacher's current level of knowledge and skill, but should also provide a foundation for subsequent professional development.
- Teacher assessments content and processes should contribute to the professionalization of teaching.

This conception of teaching and learning, and the criteria that flow from it, has strong links into psychological, educational, and measurement theory and practice. It specifies a cognitively and behaviorally complex target performance and provides a framework for examination of the impact of the assessments on the educational system of which it is a part (students, the teachers being assessed, the teaching profession, teacher educational and staff development).

Creating a Methodology for Defining Teaching Existing standards for performance assessments do not prescribe a methodology for arriving at the criteria to be used, nor should they, because there are undoubtedly many different ways to reach defensible conclusions about criteria. Unfortunately, however, although there are many general descriptions available of methods that have been used to determine content for performance assessments, few if any offer theoretically coherent rationales for the choice of criteria and assessment design methodology. As an example of how such rationales might be articulated, consider the methodology used in *The Praxis Series: Professional Assessments for Beginning Teachers™* for defining teaching through the specification of assessment criteria.

This methodology is linked to the constructivist theory of teaching and learning that was described in the previous section of this chapter. This guiding conception is explicitly linked to process for arriving at the assessment criteria, to the resultant criteria, and to the assessment methods used to collect data about these criteria. This

methodology evolved over a period of several years, and is described in detail elsewhere (Dwyer, 1993b). Figure 2-1 provides a schematic overview of this process.

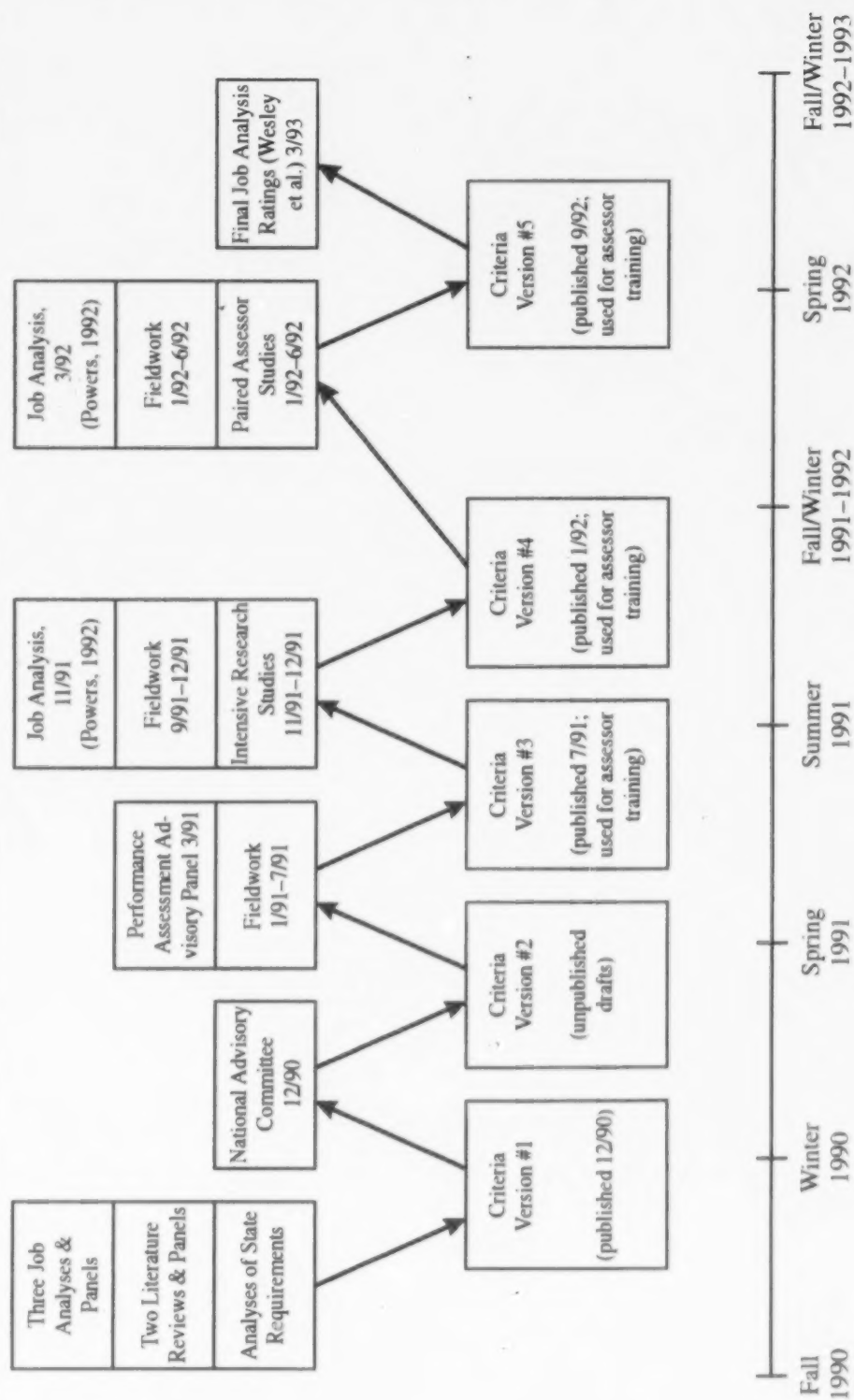
The specific job analyses, literature reviews, and state requirements, activities, and studies shown in Figure 2-1 were embedded in a larger context of field consultation and research, which is described in a later section of this chapter. Separate studies were examined for similarities and differences among the perspectives provided by the practicing educators, researchers, and the states. (The final assessment criteria for Praxis III: Classroom Performance Assessments are given in a later section of this paper.)

The creation of a theoretically-grounded methodology for arriving at assessment criteria provides evidence for performance assessment standards proposed by Frederiksen and Collins and by Linn et al., such as content quality, coverage, and transparency or meaningfulness of the criteria to the assessment participants. It also provides for the inclusion of multiple perspectives on important aspects of teaching and empirical validation of those perspectives through field work. At a pragmatic level, such a methodology also speaks to the issues of professional credibility, public acceptance, and legal defensibility.

Who Decides on Assessment Content? Once a guiding conception of teaching and learning has been articulated and a methodology specified, data bearing on the substantive aspects of teaching can be gathered and synthesized for assessment purposes. For example, standards developed and promulgated by teacher subject-matter professional organizations, state requirements for teacher knowledge and skills, student outcome standards, and research on the knowledge and skills needed by practicing teachers to perform their duties all bear on what aspects of teaching might be assessed. Although the amount of data that is amassed in a national effort to articulate assessment criteria for teaching is larger than in single statewide or local applications, a common issue presents itself in any of these situations: How to deal with the perspectives of multiple stakeholders, those who have legitimate but differing interests in the content and outcomes of the assessments. The dilemma implied by this practical situation can be simply stated: From whose perspective should the knowledge base be considered?

The three main sources of data for the Praxis III: Classroom Performance Assessments knowledge base represented three distinct perspectives: practicing teachers, educational researchers, and those who set teacher licensing requirements. The heart of this dilemma is that each of these perspectives does not simply provide a different view of the same phenomena; each asks different questions, employs differing methods to reach conclusions, and has a set of different, although often overlapping, concerns about the meaning and use of knowledge about teaching. These three views represent fundamentally different paradigms, in the sense that

Figure 2-1. Criteria Development Schematic



their basic assumptions, methodologies, and values differ. It is therefore not simply an algorithmic or mechanical process to arrive at criteria that incorporate data from these three sources. Much more useful are documentation and verification techniques borrowed from ethnography and from qualitative and naturalistic research methods in general. For example, as alluded to in the discussion of Figure 1, the developers of Praxis III: Classroom Performance Assessments resolved the dilemma by carrying out an iterative procedure of creating draft criteria derived from all three of the major data sources, then presenting the draft criteria for review to representatives of these main points of view. Reviewers and panelists were asked, in essence, if the draft criteria represented the knowledge base for teaching as they understood it. The criteria underwent a number of major revisions using this process. With each of these major revisions, increasingly large cycles of fieldwork with beginning and experienced teachers were undertaken to ensure that the resultant criteria remained meaningful to teachers themselves.

As data from the fieldwork accumulated, such practical considerations as whether those who assess the beginning teachers could understand and agree upon what was meant by a particular criterion came into play in evaluating the criteria. In addition, in order to improve teaching practice, recognizability and acceptability to the teaching profession (see Gage, 1974, for a discussion of this relationship) were explicitly used in judging the merits of the later versions of the assessment criteria. This focus on improving teaching practice is germane to establishing the consequential basis for the validity of the assessments (Messick, 1989, 1992): The assessments should contribute to the improvement of the educational system of which they are a part. It is also germane to establishing the meaningfulness of assessments to their participants.

Theoretical and Practical Knowledge

Closely related to, but distinct from the problem of resolving differing perspectives is the dilemma of dealing with the relative standing of theoretical and practical knowledge. As noted above, it is important for both practical and theoretical reasons that the criteria and their organizing framework map well to teachers' own understandings of their work. At the same time, however, the criteria should also build on educational and psychological theory, in order to give them the coherence and generalizability required by yet other validity standards, as well as an increased probability of standing the test of time in actual classroom use.

Sternberg and Wagner (1993) draw a useful distinction between academic problems and practical problems. Academic problems tend to (a) be formulated by other people than those who solve them, (b) be well defined, (c) be complete with regard to the information needed to solve them, (d) possess only a single correct

answer, (e) involve only a single method of obtaining the correct answer, (f) not be embedded in ordinary experience, and (g) be of little or no intrinsic interest. In contrast, according to Sternberg and Wagner, practical problems tend to (a) require problem recognition and formulation, (b) be ill defined, (c) require information seeking, (d) possess multiple acceptable solutions, (e) allow multiple paths to solution, (f) be embedded in and require prior everyday experience, and (g) require motivation and personal involvement to reach solutions (p. 2). The Praxis Series: Professional Assessments for Beginning Teachers™ guiding conception of teaching (Dwyer & Villegas, 1993) made it clear that in Sternberg and Wagner's terms the Praxis III: Classroom Performance Assessments assessment criteria should address the practical problems that beginning teachers must solve in order to increase their meaningfulness as well as their cognitive complexity (as distinct from difficulty).

Again, resolving the dilemma of practical and theoretical cannot be done through any simple, mechanistic process. The Praxis III solution to this problem lay once more in the more naturalistic methods of iterative reviews and revisions by the field and careful attending to diverse perspectives. Many practicing teachers and educational theoreticians reviewed and helped to revise the criteria until they were broadly perceived as acceptable from both points of view.

Lead or Lag? Very early on in the development process, determining assessment criteria involves what is often called the "lead/lag" dilemma, which raises standards issues such as fairness, utility, and legal defensibility. In the case of The Praxis Series: Professional Assessments for Beginning Teachers™, as part of the system for licensing beginning teachers, the criteria must reflect the current requirements for professional practice in order to be logically consistent with the purposes of the assessments and fair to the participants. A competing value, however, is that given the long lead time for developing high-quality assessments and the likelihood that they will continue to be used for a number of years, it is also important not to create assessments that will be, in effect, obsolete before they are completed or that will encourage continuation of teaching practices that are even now only marginally acceptable to the profession.

The crux of this dilemma is that "current requirements for professional practice" is by no means a static concept and that new knowledge about teaching is created on a daily basis. In evaluating whether a particular aspect of teaching can be considered to be supported by research or to meet current requirements for professional practice, it is thus necessary to make a number of complex judgments about the status of the research and to take into consideration the professional consensus about future trends in that area. It is also necessary to include in these deliberations some judgments about the maturity of a research area. For example, the area of teacher behavior and its links to student learning has been extensively researched

for many years. In particular subareas, the domains are well mapped, well-designed studies are numerous, and it is even possible to say that definitive conclusions have been reached.

In contrast, the area of teacher cognition and its links to student learning is still relatively young and in a state of flux. Although the importance of this research domain to teaching practice is not in dispute, its contours are still to a certain extent under discussion, and a number of important principles, although logically unsailable and convincingly demonstrated in high-quality research studies, have not yet been widely replicated, and their interconnections are not yet fully established. A complicating factor is that research on teacher behavior and research on teacher cognition tend to utilize different research methodologies, thus creating another difficulty in evaluating the newer research by traditional standards. Despite these conceptual difficulties, and despite the fact that teacher cognition has not heretofore been completely integrated into high-stakes teacher assessments, it is clear that the Praxis III: Classroom Performance Assessments would have very little credibility among teachers and other educators and researchers, now and in the future, if this perspective had been ignored. This point of view will be strengthened by standards for performance assessments that clearly require the preservation of cognitive complexity in such assessments.

Size and Scope of Criteria. Arriving at the optimal level of specificity of the assessment criteria is a difficult, iterative process that is connected to many aspects of validity including fairness, content quality, content coverage, and cognitive complexity. In general terms, this specificity issue means that if the criteria are too big, that is, too vague and general, then meaningful standards are difficult to develop and to apply fairly. People can agree in principle that a criterion represents some desirable aspect of teaching, but in practice they cannot agree on its specifics, and thus assessors cannot bring a consistent set of judgments to the assessment process. Cognitive complexity may be preserved, but only at the expense of fairness and generalizability. On the other hand, if criteria are too small, that is, too specific, people can agree on specific instances of them with great consistency, but the criteria are unlikely to be seen as capturing the essence of good teaching. In addition to failing to represent cognitive complexity, criteria that are too specific may promote a fragmented, cookbook view of teaching and thus violate another frequently-cited quality standard, improving the educational system of which they are a part.

Reaching this "just right" level of specificity is a formidable challenge, but one with wide ramifications for the assessment's technical quality. In the development of the Praxis III: Classroom Performance Assessments, achieving the level of specificity that resulted in both educationally-significant criteria and criteria that assessors could recognize in specific instances involved many iterations of field-

Table 2-2. Praxis III: Classroom Performance Assessment Criteria

Domain A. Organizing Content Knowledge for Student Learning

- A1: Becoming familiar with relevant aspects of students' background knowledge and experiences
- A2: Articulating clear learning goals for the lesson that are appropriate for the students
- A3: Demonstrating an understanding of the connections between the content that was learned previously, the current content, and the content that remains to be learned in the future
- A4: Creating or selecting teaching methods, learning activities, and instructional materials or other resources that are appropriate for the students and that are aligned with the goals of the lesson
- A5: Creating or selecting evaluation strategies that are appropriate for the students and that aligned with the goals of the lesson

Domain B. Creating an Environment for Student Learning

- B1: Creating a climate that promotes fairness
- B2: Establishing and maintaining rapport with students
- B3: Communicating challenging learning expectations to each student
- B4: Establishing and maintaining consistent standards of classroom behavior
- B5: Making the physical environment as safe and conducive to learning as possible

Domain C. Teaching for Student Learning

- C1: Making learning goals and instructional procedures clear to students
- C2: Making content comprehensible to students
- C3: Encouraging students to extend their thinking
- C4: Monitoring students' understanding of content through a variety of means, providing feedback to students to assist learning, and adjusting learning activities as the situation demands
- C5: Using instructional time effectively

Domain D. Teacher Professionalism

- D1: Reflecting on the extent to which the learning goals were met
- D2: Demonstrating a sense of efficacy
- D3: Building professional relationships with colleagues to share teaching insights and to coordinate learning activities for students
- D4: Communicating with parents or guardians about student learning

work analysis and revision of the criteria. The fieldwork for the Praxis III: Classroom Performance Assessments was designed to try out draft criteria in a number of settings (different types of subject matter, schools, students, age levels, etc.) and to evaluate the criteria and procedures for collecting data about them from a number of perspectives. In the reports of this fieldwork (Myford et al., 1993, provide an overview of it), there are numerous instances of these experiences leading the developers to conclude, for example, that what had been a single criterion ought to be divided into two separate criteria to help assessors better understand how a particular aspect of teaching is actually played out in the classroom and help them recognize evidence related to this aspect of teaching when they see it.

In other instances, researchers concluded that particular criteria were seen differently by teachers of particular subject matters or age groups. As noted above organizing and wording the criteria so that they are clear and logical from the point of view of those who use them was given a high priority in the development work, both for reasons of improving teaching practice as a result of participating in the assessment process and for considerations of content quality, coverage, and meaningfulness to teachers. The final criteria for the Praxis III: Classroom Performance Assessments are given in Table 2-2. Note that, consistent with the guiding conception of teaching, these represent salient aspects of teaching, not particular behaviors. That is, they serve as a framework for the assessors by representing *what* proficient teachers attend to rather than *how* they implement these aspects, which is highly context-sensitive. Moreover, the criteria, as representations of complex performances, are not intended to be construed as independent constructs, but as facets of a single construct. As noted above, assessment of the underlying pedagogical and discipline knowledge that enables acceptable performance on these criteria is carried out in other parts of The Praxis Series: Professional Assessments for Beginning Teachers™ and thus not a part of the set of criteria shown in Table 2-2.

The results of this process offer evidence for many of the standards proposed for performance assessments that were discussed in an earlier section of this paper. For example, Claxton, Murrell, and Porter's proposed standards (1987) dealing with consistency of educational goals, feasibility, assessor involvement in assessment development and administration, and perceived utility to the assessee are all issues addressed directly in criterion development field work (Myford et al., 1993). Similarly, this fieldwork provided evidence bearing on Linn, Baker, and Dunbar's (1991) proposed standards for fairness (in terms of teaching area and preferred style, as well as in terms of race, ethnicity, gender, etc.), cognitive complexity, content quality and coverage, meaningfulness, and efficiency. It will be important in this and other projects to begin mapping these linkages across standards in order to amass a knowledge base of approaches to providing evidence for these standards and suggestions for refinement of the proposed standards.

Professional Judgment in Teacher Performance Assessments. The specification of assessment criteria is an important part of developing any performance assessment, but the success of the effort as a whole can only be evaluated in light of the ability of assessors to use these criteria to reach technically and professionally defensible conclusions. Unlike traditional multiple-choice testing, where the great majority of the professional judgment comes into play during the preadministration phase of test development, professional judgment in performance assessments is required in both the development and the use phases of the assessment. The quality of this professional judgment impacts many aspects of the assessment's validity, including, but not limited to, fairness, cognitive complexity, and construct representation. It is also related to concerns for generalizability, although the concerns are not the same for classroom performance assessment as they are for simulation-based assessment. In classroom performance assessment, the generalizability concern is to other teaching events, not to other aspects of teaching. In this sense, the teaching events are analogous to exercises or tasks in other types of performance assessments; in the Praxis III: Classroom Performance Assessments model, the "scoring" of the "tasks" is held constant via the criteria and their associated scoring rules.

In assessing live teaching performance, variability across "tasks" is a natural and acceptable phenomenon, and thus inferences based on a given set of teaching events are expected to generalize to an intrinsically variable universe of teaching events that defines the construct. Generalizability across tasks is therefore not problematic in the same sense as when tasks are seen as partial or indirect instantiations of the constructs (for example, in the work of Shavelson and his colleagues on student performance assessments in science [Shavelson, Baxter, & Pine, 1991]). As noted above, the Praxis III: Classroom Performance Assessments criteria are intended to be construed as interrelated aspects of a complex performance, not as functionally independent entities. As such, one would not aim to generalize from one aspect of teaching to another as evidence of validity, but rather to investigate the patterns of ratings given across occasions and within a single occasion by two or more assessors (assuming that occasions are expected to be highly variable, relative to within- or across- assessor variability).

In the Praxis III: Classroom Performance Assessments, trained assessors gather data about the assessment criteria using a variety of interrelated methods: interviews with the beginning teacher, observation in the teacher's own classroom, and written documents about the students and the learning that is to take place. Given their responsibilities, assessors' training is necessarily extensive. In addition to providing opportunities for the assessors to build meaning for the criteria, assessor training covers such topics as observation and interviewing skills, assessor ethics, defensible documentation, and the ability to recognize evidence of the criteria in a

variety of contexts and teaching styles that may or may not match the assessor's personal experiences and preferences.

As noted above, the criteria do not stand alone; because they are aspects of teaching, not particular behaviors, they must be interpreted in light of the actual classroom context, which includes both the students and the subject-matter being taught. The criteria serve as the guide for structuring assessors' judgments, ensuring that a common frame of reference rather than personal preference is the basis of the assessors' conclusions and ratings. Assessor judgment is thus the cornerstone of the defensibility of the ratings of the beginning teacher. Using the methods described above, assessors gather and organize data bearing on each of the criteria; make critical judgments about the importance of the evidence and its relevance to particular criteria; then reach a conclusion about the beginning teacher's level of performance on each criterion based on this evidence and their interpretation of it.

Assessors document these judgments by citing specific evidence and linking it to a rating scale that describes increasingly proficient levels of performance with respect to each of the criteria. Legitimacy of the assessment process is thus based on the quality of this argumentation (structured, documented, professional judgment) rather than on a purported absence of human decision-making (objectivity). In this way, important aspects of validity can be accommodated in the assessment process, such as directness of measurement, context-sensitivity, and adequacy of construct representation. Through special studies (such as paired-assessor comparisons), field work in a variety of teaching settings, and operational use, various methods of data gathering may be found to result in better measurement—that is, in more accurate or detailed judgments of the criteria, in better documentation, or in more positive effects on the system of which the assessment is a part. The data-gathering methods themselves, however, are clearly subordinate to the quality of the criteria and the assessors' judgments in determining the value and validity of the assessments.

Conclusion

Validity theory, together with currently-available and emerging standards for performance assessments provide guidance for the developers of high-stakes performance assessments. It is imperative, however, that important aspects of validity and standards for quality and fairness of performance assessments be built into such assessments from their very inception. Specifying the target performance in terms legitimate to all of the assessment participants and creating an explicit methodology for integrating diverse points of view provide the foundation for defensible assessments. It is only through painstaking analyses and field work, however, that many validity-related aspects of the assessments can be satisfactorily

resolved. Perhaps, with the passage of time, a cycle can be established in which these experiences from the field can inform further development of standards for performance assessment, which can then be used to raise the standard of assessment development practice. Only then can the full promise of modern validity theory be fulfilled.

References

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Educational Association.
- Claxton, C., Murrell, P. H., & Porter, M. (1987). Outcomes assessment. *AGB Reports*, 29(5), 32-35.
- Cole, N. S., & Moss, P. A. (1989). Bias in test use. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 201-219). New York: Macmillan.
- Dwyer, C. A. (1992). Classroom observations for licensing beginning teachers. In Educational Testing Service, *What we can learn from performance assessment in the professions: Proceedings of the 1992 ETS Invitational Conference*. Princeton, NJ: Educational Testing Service.
- Dwyer, C. A. (1993a). Teaching and diversity: Meeting the challenges for innovative teacher assessments. *Journal of teacher education*, 44, 119-129.
- Dwyer, C. A. (1993b). *Development of the knowledge base for the Praxis III: Classroom Performance Assessments Assessment Criteria*. Princeton, NJ: Educational Testing Service.
- Dwyer, C. A., & Villegas, A. M. (1993). *Guiding conceptions and assessment principles for The Praxis Series: Professional Assessments for Beginning Teachers™*. Princeton, NJ: Educational Testing Service.
- Educational Testing Service. (1992). *Guidelines for proper use of The Praxis Series: Professional assessments for beginning teachers*. Princeton, NJ: Author.
- Educational Testing Service. (1993). *Using the Praxis III: Classroom Performance Assessments for teacher licensing*. Princeton, NJ: Author.
- Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, 18(9), 27-32.
- Gage, N. L. (1974). Evaluating ways to help teachers to behave desirably. In *Competency assessment research and evaluation: A report of a national conference, March 12-15, 1974*. Houston, TX: University of Houston.
- Haertel, E. H. (1991). New forms of teacher assessment. In G. Grant (Ed.) *Review of Research in Education* (Volume 17, pp. 3-29). Washington, DC: American Educational Research Association.

- Joint Committee on Standards for Educational Evaluation. (1988). *The personnel evaluation standards*. Newbury Park, CA: Sage.
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20(8), 15-21.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.
- Messick, S. (1992, April). The interplay of evidence and consequences in the validation of performance assessments. Invited address to the annual meeting of the National Council on Measurement in Education, San Francisco.
- Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research*, 62, 229-258.
- Miller, M. D., & Legg, S. M. (1993). Alternative assessment in a high-stakes environment. *Educational Measurement: Issues and Practice*, 12(2), 9-15.
- Myford, C., Villegas, A. M., Reynolds, A., Camp, R., Jones, J., Knapp, J., Mandinach, E., Morris, L., & Sjoström, B. (1993). *Formative studies of Praxis III: Classroom Performance Assessments, an overview*. Princeton, NJ: Educational Testing Service.
- National Board for Professional Teaching Standards (1991). *Toward high and rigorous standards for the teaching profession* (3rd ed.). Detroit, MI: Author.
- Quellmalz, E. S. (1991). Developing criteria for performance assessments: The missing link. *Applied Measurement in Education*, 4, 319-331.
- Scriven, M. (1990). Can research-based teacher evaluation be saved? *Journal of Personnel Evaluation in Education*, 4, 19-32.
- Shavelson, R. J., Baxter, G. P., & Pine, J. (1991). Performance assessment in science. *Applied Measurement in Education*, 4, 347-362.
- Sternberg, R. J., & Wagner, R. K. (1993). The g-centric view of intelligence and job performance is wrong. *Current Directions in Psychological Science*, 2(1), 1-4.
- Tittle, C. K. (1989). Validity: Whose construction is it in the teaching and learning context? *Educational Measurement: Issues and Practice*, 8, 5-13, 34.

3

SCHOOL PROFESSIONALS' GUIDE TO IMPROVING TEACHER EVALUATION SYSTEMS

Preamble: A Guide to Improving Teacher Evaluation Systems

Part 3 is a step-by-step **GUIDE** to assessing and improving teacher evaluation systems. The **GUIDE** demonstrates how the *Standards* can be effectively and systematically operationalized to examine presently used, or contemplated, teacher evaluation systems. Core duties to be considered in evaluating a teacher's performance are more closely defined as the normative list of what teachers legitimately can be held responsible for (Scriven, 1988). Discussion then moves to a consideration of which particular Joint Committee *Standards* should be used to assess the adequacy of a teacher evaluation system, and this is followed by the presentation of a conceptual framework delineating factors that define and influence performance evaluation systems. This framework is applicable to a range of personnel evaluation systems in education. Based on this framework, schools and school districts are invited to examine personnel evaluation systems step by step. The term **GUIDE** is seen to be most appropriate in this regard.

Advice is then offered about how to organize a participatory project to improve teacher evaluation. The politics of such a situation, including importantly the involvement of all stakeholder groups, are then addressed, together with activities to inform and convince the district community that change is essential.

Finally, steps are outlined for formulating a new system for teacher evaluation or considering an improved version of an existing one. This initially entails determining which personnel evaluation *Standards* are met. A team effort is advised to achieve this significantly important objective. This somewhat arduous, but essential, process is considerably helped by the GUIDE's clearly depicted steps, supported by strong reference to the *Standards*, an inventory form for documenting a teacher evaluation system (given in Appendix A), and a series of questions to be answered in addressing each standard when extant systems are being examined (in Appendix B).

A version¹ of a GUIDE for all school-connected personnel, prepared by Bernard McKenna, David Nevo, Daniel Stufflebeam (Project Director), and Rebecca Thomas under the auspices of the federally funded Teacher Evaluation Improvement Program of the Center for Research on Educational Accountability and Teacher Evaluation (CREATE)

Introduction

Competent, dedicated, and well-performing teachers are any school's most important resource. Teachers are the professionals most directly responsible for helping all students to learn, and students benefit or suffer from the quality of the teaching they receive. Moreover, any society is at risk when its schools fail to educate its children and youth. So, clearly, effective teaching must be assured; and the teaching profession, school boards, school administrators, and school faculties must recognize that teacher evaluation is a key means of providing that assurance.

Decisions in selecting teachers (though not the focus of this GUIDE) should be informed by sound evaluations of the candidates. Without that protection, a school district is unlikely to succeed in a number of its important missions. But teacher selection evaluations alone are not enough. They predict, but cannot guarantee. Systematic evaluation of teachers' performance in schools—the focus of the GUIDE—is essential.

Schools need to and often do hedge selection decisions by placing new teachers on probation, during which time they are expected to demonstrate competence and effectiveness. Systematic evaluation during the probationary period can be especially useful, since it provides the new teachers with feedback for improving performance, and the school district with an informed basis for deciding whether

¹ The complete GUIDE is available from the National Center for Research on Educational Accountability and Teacher Evaluation (CREATE), The Evaluation Center, Western Michigan University, Kalamazoo MI, 49008-5178.

or not to extend a teacher's contract. However, decisions to extend the contract or even to award tenure do not exhaust the valuable uses of teacher evaluation. Extended and tenured teachers must be periodically evaluated to provide them with feedback for examining and strengthening their service. In addition, the school district needs to evaluate the performance of *all* its teachers as a sound basis for remediating or terminating those few who become persistently poor performers and for recognizing and reinforcing outstanding teaching.

In response to these pervasive needs for sound teacher evaluations, virtually all U.S. school districts implement some type of evaluation system for informing teacher selection decisions and evaluating on-the-job performance. However, as Scriven, Wheeler, and Haertel (1992, 1993) and others have documented, schools often are dissatisfied with their teacher evaluation systems. And for good reason. Teacher evaluations in use often are

- not grounded in clear rationale and policy
- not focused on defensible criteria
- not reliable
- not credible
- not sensitive to particular teaching settings
- not influential
- biased
- superficial
- demoralizing

It is not surprising then that many school districts are seeking assistance in assessing and improving their teacher evaluation systems. Their teacher selection and supervision activities are heavily dependent on teacher evaluation, but the evaluation results, though costly in time and resources, often are not professionally defensible or satisfactory to anyone. Thus, despite good intentions, students and teachers may not be well served.

This GUIDE is directed toward helping school districts to assess and improve their teacher evaluation systems, i.e., the evaluations they use to assess the performance of probationary, extended, and tenured teachers. It is a GUIDE for school professionals, school board members, consultants, parents, students, and other stakeholders to use in documenting and examining their current teacher evaluation system and planning and making needed improvements.

The GUIDE is grounded in professional standards that define and describe sound teacher evaluations. By using the *Standards* a school district can determine in what important respects its teacher evaluation system is succeeding or failing. The GUIDE is also keyed to the findings of research on existing teacher evaluation systems being conducted by the national Center for Research on Educational

Accountability and Teacher Evaluation (CREATE). This research² was used to identify the full range of specific variables that can create a defective or substandard teacher evaluation system.

The professional standards for sound teacher evaluations and the research on teacher evaluation systems used to develop this GUIDE are complementary. *The Standards provide the basic criteria for determining whether a teacher evaluation system is satisfactory in concept, design, operation, and outcomes. The research on teacher evaluation systems provides detailed direction for looking closely at identified deficiencies and diagnosing and correcting them.*

The underlying strategy in compiling the GUIDE is to provide a step-by-step process for examining and improving a teacher evaluation system. Basically, this process includes the following steps:

- Develop and adopt a guiding philosophy and concept of teacher evaluation
- Provide a framework for involving all interested stakeholders in the process of examining and improving the district's teacher evaluation system
- Inventory and carefully describe the district's current teacher evaluation system
- Judge the current teacher evaluation system against the Joint Committee Personnel Evaluation Standards
- Diagnose the particular issues and problems that must be addressed in improving the teacher evaluation system
- Redesign the system
- Develop and obtain support for a project to install and implement the improved teacher evaluation system

This GUIDE is especially intended for the use of teacher evaluation improvement teams in school districts. Such teams, or committees, should include representatives of all groups involved in or affected by the district's teacher evaluation system. For example, the improvement team might include

- school board member
- superintendent or assistant superintendent
- director of personnel or other personnel office staff
- elementary school principal
- middle school principal

² More details on developing a grounded theory for teacher evaluation can be found in the following internal CREATE documents: Stufflebeam, D. L., & Nevo, D. (1992). Toward a theory of teacher evaluation and Stufflebeam, D. L., & Thomas, R. (1993). Evaluation theory: Development of a grounded theory of teacher evaluation.

- secondary school principal
- elementary school parent
- middle school parent
- secondary school parent
- elementary school teacher
- middle school teacher
- secondary school teacher
- middle school student
- secondary school student
- school psychologist
- counselor
- school research and evaluation specialist

Such a team should find that this GUIDE provides clear, practical advice about what steps to take in examining and improving their system. The GUIDE is presented in straightforward language and includes forms on which to organize key information and record the team's decisions. It is also designed to be useful to parents, school board members, and others who have a legitimate interest and role in improving the teacher evaluation system.

This GUIDE is a companion document to the Joint Committee's (1988) *Personnel Evaluation Standards*. Users of the GUIDE are advised to obtain *The Personnel Evaluation Standards* and to use the two documents in combination. Together, they provide a sound basis for examining and redesigning teacher evaluation systems.

Teacher Evaluation: Its Purpose, Meaning, and Improvement

Teacher evaluation is a pervasive concern of community members and school district personnel. Employers, school staff, students, parents, and others share an interest in assuring effective teaching in their school districts, and to help that happen they support sound teacher evaluation practices. So that these stakeholders can assess and assist school district efforts, districts' needs to adopt and communicate a sound, clear concept of teacher evaluation.

The purpose of this section is to present, for consideration by school districts, a state-of-the-art concept of what teacher evaluation is, what it should be, and what is involved in improving it. To define the foundation principles of sound teacher evaluation, this conceptualization draws particularly from the work of the national Joint Committee on Standards for Educational Evaluation.

In order to consider the full range of relevant practical issues, this section employs research on teacher evaluation practices being conducted by the federally-

funded national Center for Research on Educational Accountability and Teacher Evaluation (CREATE). It draws especially on a CREATE project that used the methodology of grounded theory development to identify, analyze, and synthesize the full range of variables that contribute to the success or failure of teacher evaluation systems³. The grounded theory project based its findings on in-depth study of a range of actual teacher evaluation systems. An additional source is the CREATE Teacher Evaluation Models Project which proposes a definition of the generic duties of teachers and recommends that these be used as the basis for evaluating teacher performance (Scriven, Wheeler, & Haertel, 1992 and 1993; Scriven, 1994).

This section is presented not as a philosophical and theoretical treatise, but as a straightforward response to key questions concerning a school district's philosophy and mission of teacher evaluation.

1. *What is Sound Teacher Evaluation?*

Consistent with the definition of personnel evaluation provided by the Joint Committee on Standards for Educational Evaluation (1988), **teacher evaluation** is defined here as the *systematic assessment of a teacher's performance and/or qualifications in relation to the teacher's defined professional role and the school district's mission*.

It is important to note that this definition calls for **systematic** assessment. It does not sanction haphazard, casual exercises that often masquerade as sound evaluation. It also requires that teachers be assessed for their effectiveness in carrying out their **defined assignments** and their contribution to fulfillment of the district's mission and not for their personalities and particular styles of teaching. Proper use of this definition requires that districts clearly define their mission and the roles of individual teachers in carrying it out, and then use these as the basic criteria for evaluating the teacher's performance.

2. *What Core Duties Should Be Considered in Evaluating a Teacher's Performance?*

Michael Scriven (1988, 1990) has warned against the pitfalls of evaluating teachers on factors identified from correlational research. Those teacher characteristics and behaviors that might correlate best with student learning, such as gender, race,

³ See footnote 2 in previous section.

physical handicaps, and similar criteria, cannot, according to law and ethics, be used to make personnel decisions. Other teacher variables that might show a misleading high correlation with student outcome measures include styles of teaching, e.g., injecting humor, showing enthusiasm, using an inquiry approach, or issuing punishments and rewards.

Scriven argues that it is absolutely inappropriate and invalid to evaluate a teacher on criteria selected only because they show moderate to high correlations with student achievement measures. The correlations are never perfect and usually not even high. Moreover, applying such criteria places a greater value on the variable than on the desired result; it would penalize the teacher who scores low on the predictor measure but who nevertheless is effective in helping students to learn.

Also, the omission of criteria previously shown to have low correlations with student achievement measures might deemphasize some critical teacher responsibilities, such as knowledge of course content, ability to communicate course content clearly to students, ability to manage classroom activities, ability to examine student progress, and treating students fairly and equitably. Irrespective of research-supported correlations with student outcome measures, the importance of these responsibilities in teaching is universally acknowledged.

As an alternative to the popular approach of basing the selection of teacher evaluation criteria on the results of correlational research, Scriven has recommended that teachers be evaluated directly on their fulfillment of duties. For Scriven, the **core duties**, around which other duties are defined, are *the normative list of what teachers can legitimately be held responsible for knowing and doing*. His recommendation is consistent with the tradition in personnel psychology requiring that performance be evaluated in terms of job descriptions. But it goes further. It calls for the teaching profession to be clear about the pervasive ethical responsibilities of teachers, wherever they may serve, and it calls on school districts to specify the list of teaching duties in their description of each teacher's responsibilities and to focus on them in evaluations of teacher performance.

Through many years of interaction with teachers and school administrators, Scriven (1994) has evolved the following list of core duties for use in evaluating teacher competence and performance:

1. Knowledge of subject matter
 - Field of special competence
 - Pervasive curriculum subjects
2. Instructional competence
 - Communication skills
 - Classroom management
 - Course development
 - Course evaluation

3. Assessment
 - Testing
 - Grading
 - Reporting
4. Professionalism
 - Ethics
 - Attitudes
 - Service
 - Knowledge of duties
 - Knowledge of school and its context
5. Other individualized services to the school and community

We include the list here as a guide for school district personnel as they examine and improve their teacher evaluation systems. Scriven's rationale for and definitions of the core duties are discussed extensively in his papers referenced at the conclusion of the chapter.

3. What Might Be Included in the Rationale for Evaluating Teacher Performance?

It is important that all stakeholders in a school district have in mind a common, clear, and defensible rationale for evaluating teacher performance. Some of the compelling reasons for teacher evaluation are to

- Foster high quality service to students
- Help teachers to assess and improve competence
- Motivate and assist teachers to constantly assess and improve instruction
- Maintain teacher accountability
- Recognize and reward outstanding teaching
- Identify and remediate ineffective teaching
- Safeguard student and community interests from incompetent or harmful teaching
- Terminate persistently poor teachers
- Oversee and coordinate teaching across classrooms
- Assess teaching performance as a basis for planning professional development
- Enhance school credibility

4. *What Is a Standard for Sound Teacher Evaluation?*

The Joint Committee on Standards for Educational Evaluation (Joint Committee, 1988) defined a standard as "*a principle commonly agreed to by people engaged in the professional practice of evaluation for the measurement of the value or the quality of an evaluation*" (p. 187).

All persons involved in or affected by a school district's teacher evaluation practices have a right to expect that teacher evaluations are designed and carried out in full compliance with the established professional standards and requirements. Thus, teacher evaluations themselves are subject to evaluation, and the foundation for assessing them is the published professional standards for judging evaluations of teachers and other educational personnel.

5. *Why are professional standards for evaluations important?*

In any field that provides professional service to the public—such as medicine, law, accounting, auditing, psychiatry, engineering, and teaching—the professionals must live by their profession's standards of sound and ethical practice. The standards are determined and periodically updated by representatives of the profession. And the standard-setting process includes input from research on practice, examination of the quality and positive and negative outcomes of past practice, review of relevant court cases, and feedback from clients and other stakeholders—all thoroughly processed to reach professional consensus. Adherence to the standards is intended to provide clients with high quality service and to protect them from the harmful effects of substandard or unethical practice.

Thus, standards for sound teacher evaluation protect students, teachers, and others from incompetent or misguided teacher evaluation practices, and help to assure that evaluations lead to improved teaching through sound feedback to teachers and their supervisors.

6. *What Particular Standards Should Be Used to Judge the Adequacy of a Teacher Evaluation System?*

The Joint Committee on Standards for Educational Evaluation used a national consensus process to establish and articulate as a fundamental proposition that all educator evaluation systems should have four basic attributes: *propriety, utility, feasibility, and accuracy*.

In order to articulate each of these attributes, the Joint Committee defined 21 specific standards and explicated each with practical guidelines, common errors to be avoided, and illustrations of application. The set of standards is designed to help educators examine the extent to which any personnel evaluation system possesses the four essential attributes, identify system deficiencies to be corrected, and/or develop appropriate, effective new systems.

Following are the Committee's definitions of the four basic attributes of a sound personnel evaluation, followed by summary listings of the applicable standards that enhance each of them. The specific guidelines, common errors to be avoided, and illustrative cases for each standard appear in the original publication of the Standards (Joint Committee, 1988)⁴.

To emphasize the direct applicability of these Standards to teacher evaluation, the word "teacher" has been substituted wherever the Joint Committee Standards use the words "personnel" or "educator." The only intention of this modification is to focus discussion in the GUIDE on evaluations of the performance of teachers, rather than on educational personnel generally. School districts, of course, should evaluate the performance of all their personnel and *The Personnel Evaluation Standards*, enumerated below, are equally applicable to assessments of the performance of administrators, counselors, librarians, and all other school personnel.

Propriety is aimed at protecting the rights of students, teachers, administrators, evaluators, and other persons affected by an evaluation system. The inclusion of propriety standards reflects the fact that teacher evaluations may violate or fail to address certain ethical and legal principles. The primary principle is that schools exist to serve students; therefore, teacher evaluations must concentrate on determining whether teachers are effectively meeting the educational needs of students. Overall, the Propriety Standards require that evaluations be conducted legally, ethically, and with due regard for the welfare of students, teachers, and other involved or affected parties.

In order to satisfy the condition of propriety, teacher evaluations should adhere to the following Joint Committee standards:

P1: Service Orientation

Evaluations of teachers should promote sound education principles, fulfillment of institutional missions, and effective performance of job responsibilities, so that the educational needs of students, community, and society are met.

⁴ Appreciation is expressed to the Joint Committee on Standards for Educational Evaluation for giving its permission to reprint the summary statements of the 21 Personnel Evaluation Standards.

P2: Formal Evaluation Guidelines

Guidelines for teacher evaluations should be recorded and provided to employees in statements of policy, negotiated agreements, and/or teacher evaluation manuals, so that evaluations are consistent, equitable, and in accordance with pertinent laws and ethical codes.

P3: Conflict of Interest

Conflicts of interest should be identified and dealt with openly and honestly, so that they do not compromise the evaluation process and results.

P4: Access to Personnel Evaluation Reports

Access to reports of teacher evaluation should be limited to individuals with a legitimate need to review and use the reports, so that appropriate use of the information is assured.

P5: Interactions with Evaluatees

The evaluation should address teachers in a professional, considerate, and courteous manner, so that their self-esteem, motivation, professional reputations, performance, and attitude toward personnel evaluation are enhanced or, at least, not needlessly damaged.

Utility is intended to make evaluations informative, timely, and influential. Especially, it requires that evaluations provide information useful to individual teachers and to groups of teachers in improving their performance. Utility also requires that evaluations be focused on predetermined uses, such as informing selection and promotion decisions or providing direction for staff development, and that they be conducted by persons with appropriate expertise and credibility. In general, teacher evaluation is viewed as an integral part of an institution's ongoing effort to recruit outstanding teachers and, through timely and relevant evaluative feedback, to encourage and guide them to deliver high quality service.

Utility standards should be especially welcome to teachers who see their institution's performance review system as only ritualistic and not helpful or, worse, demoralizing and counterproductive. By applying the Utility Standards, an institution is guided to clarify intended uses of its evaluation system and of particular teacher evaluations, and to do whatever is necessary to ensure that the system addresses relevant questions, issues useful feedback, provides direction for improvement, and does not decide on uses after the fact. The main point of the Utility Standards is to insure that evaluations contribute constructively to helping teachers and other educators deliver excellent service.

Standards that enhance the utility of an evaluation are:

- U1: Constructive Orientation**
Evaluations should be constructive, so that they help institutions to develop human resources and encourage and assist teachers and other educators to provide excellent service.
- U2: Defined Uses**
The users and the intended uses of a teacher evaluation should be identified, so that the evaluation can address appropriate questions.
- U3: Evaluator Credibility**
The evaluation system should be managed and executed by persons with the necessary qualifications, skills, and authority, and evaluators should conduct themselves professionally, so that the evaluation reports are respected and used.
- U4: Functional Reporting**
Reports should be clear, timely, accurate, and germane, so that they are of practical value to the teacher and other appropriate audiences.
- U5: Follow-Up and Impact**
Evaluations should be followed up, so that users and teachers are aided to understand the results and take appropriate actions.

Feasibility emphasizes the reality that teacher evaluations (or other personnel evaluations) are conducted in institutional settings that have limited resources and instructional time and are influenced by a variety of social, political, and governmental forces. Accordingly, the Feasibility Standards call for evaluation systems that are efficient, easy to use, not disruptive of the teaching/learning process, adequately funded, and politically viable.

The Feasibility Standards are listed below:

- F1: Practical Procedures**
Teacher evaluation procedures should be planned and conducted, so that they produce needed information while minimizing disruption and cost.
- F2: Political Viability**
The teacher evaluation system should be developed and monitored collaboratively, so that all concerned parties are constructively involved in making the system work.

F3: Fiscal Viability

Adequate time and resources should be provided for teacher evaluation activities, so that evaluation plans can be effectively and efficiently implemented.

Accuracy, the fourth requirement, emphasizes the need to determine whether an evaluation has produced dependable information about relevant qualifications or performance of a teacher or other educator. This requires that the information obtained be technically defensible and that the conclusions be linked logically to the data. The position underlying the accuracy standards is that performance criteria must be derived from a valid description of the teacher's job. Simply showing that a personal characteristic—such as management style, quantitative aptitude, or race—is correlated with student achievement is not justification for using that characteristic to measure and judge a teacher. As Scriven (1988) has argued, to do so not only risks prejudicial treatment of individuals but, since the correlations are based on group data and are never perfect, such practice also produces invalid assessments of teachers who rate low on the variable but teach well, or rate high and are ineffective. Our field tests clearly indicated that many teacher evaluation systems need to be improved in how well they define jobs, how realistically they consider environmental influences, how validly they measure job qualifications and performance, and how effectively they control for various kinds of bias.

To assess the accuracy of a personnel evaluation, the Joint Committee presented the following eight Accuracy Standards:

A1: Defined Role

The role, responsibilities, performance objectives, and needed qualifications of the teacher should be clearly defined, so that the evaluator can determine valid assessment criteria.

A2: Work Environment

The context in which the teacher works should be identified, described, and recorded, so that environmental influences and constraints on performance can be considered in the evaluation.

A3: Documentation of Procedures

The evaluation procedures actually followed should be documented, so that the teachers and other users can assess the actual, in relation to intended, procedures.

A4: Valid Measurement

The measurement procedures should be chosen or developed and implemented on the basis of the described role and the intended use, so that the inferences concerning the teacher are valid and accurate.

A5: Reliable Measurement

Measurement procedures should be chosen or developed to assure reliability, so that the information obtained will provide consistent indications of the performance of the teacher.

A6: Systematic Data Control

The information used in the evaluation should be kept secure, and should be carefully processed and maintained, so as to ensure that the data maintained and analyzed are the same as the data collected.

A7: Bias Control

The evaluation process should provide safeguards against bias, so that the teacher's qualifications or performance are assessed fairly.

A8: Monitoring Evaluation Systems

The teacher evaluation system should be reviewed periodically and systematically, so that appropriate revisions can be made.

The Personnel Evaluation Standards provide a comprehensive and widely endorsed basis for assessing and improving teacher evaluation systems. They were developed by the major professional organizations that represent the full range of professionals who work in school districts, including, among others, the American Evaluation Association, the American Educational Research Association, the American Federation of Teachers, the National Education Association, the National School Boards Association, the American Association of School Administrators, and the Association for Supervision and Curriculum Development. Teachers, administrators, board members, and other education stakeholders all stand to benefit through use of the *Standards* as a tool for examining and improving teacher evaluations.

7. How Can a District Learn to Use the Personnel Evaluation Standards?

Understanding the standards is the first step in any systematic attempt to use them to develop or improve a particular teacher evaluation system. All those involved in

assessing and improving a teacher evaluation system should gain a working knowledge of the *Standards*. Given their low cost and their presentation in clear lay language, a school district will find it beneficial and feasible to make copies available to their board members and school staff. It should be borne in mind that *The Personnel Evaluation Standards* are designed to examine the full range of selection and performance evaluation systems used for the full range of professional educators. Thus, a school district will find them useful for much more than only assessing and strengthening the teacher evaluation system. The next section provides advice on how to help school professionals and others to develop a working knowledge of *The Personnel Evaluation Standards* and how to apply them.

8. If a Teacher Evaluation System Is Deficient in Meeting the Personnel Evaluation Standards, How Can a School District Team Find Out What Specifically Needs to Be Done to Improve the System? In Other Words, What Variables Should Be Considered In Revising a Teacher Evaluation System So That It Meets All the Joint Committee Standards?

Once a school district determines which Joint Committee standards are met and which are not met by its teacher evaluation system, it needs to diagnose the problems to be solved and the strengths to be preserved in order to improve the system. The *Standards* have been constructed with these needs in mind.

Careful examination of *The Personnel Evaluation Standards* reveals that they are delineated in several layers of abstraction. Going from the more abstract to the more concrete, we see at the first layer the fundamental proposition that all evaluations should have the four *basic attributes* already discussed: propriety, utility, feasibility, and accuracy. At the second layer are the 21 *Standards* listed above, that if met will assure that the evaluation has the above-mentioned four basic attributes. At the third layer are the *guidelines* for each standard, which provide procedural suggestions intended to help meet the requirements of each standard plus *common errors* to be avoided. At the fourth layer are the *illustrative cases*, concrete examples of how each standard could actually be applied.

The common errors listed for each standard provide a useful starting point for identifying deficiencies that need to be corrected, and the guidelines give useful ideas for corrective action. But experience in applying the Standards shows that the common errors and guidelines for each standard are only the beginning of the diagnostic/prescriptive process.

Therefore, CREATE researchers undertook a systematic study of actual teacher evaluation systems. The issue of what would be a more comprehensive set of variables that might interactively determine the quality of a teacher evaluation

system, as measured against the 21 standards was closely examined. It was found that an extensive array of such variables that potentially need to be taken into account in any district's efforts to assess and improve its evaluation system exists. These variables include virtually all the standards, guidelines, and pitfalls enumerated in *The Personnel Evaluation Standards*, which reinforce the validity of the Standards. But we also found many variables not explicitly included in the standards. Thus, the results of our study can extend and enrich the material in the Joint Committee Standards.

We synthesized the identified variables into a general conceptual scheme (see Figure 3-1). This synthesis aims at providing evaluation system improvement teams with a convenient overview of the variables involved in the workings of a teacher evaluation system and portraying how major groupings of these variables interact to determine the quality of teacher evaluations. This scheme summarizes, categorizes, and shows general interrelationships among the full set of identified variables for defining a given teacher evaluation system and determining its strengths and weaknesses.

The variables are divided into *context*, *inputs*, *processes*, and *products* that interact to cause and manifest the success or failure of a teacher evaluation system.

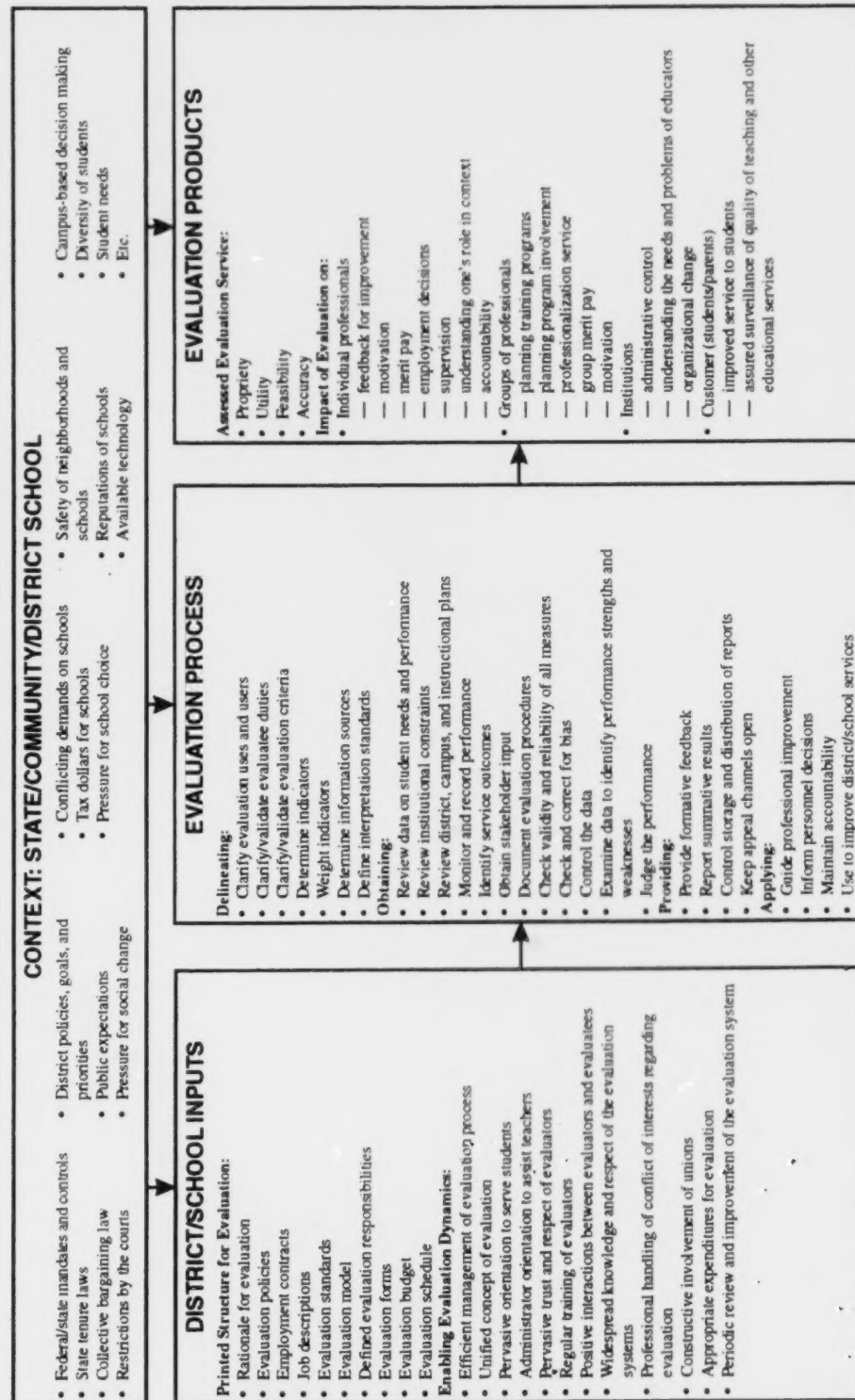
One set of variables which influences all inputs, processes and products, relates to the *context* in which the evaluation occurs. As Figure 3-1 shows, the *context* in which the school district functions includes state, community, school district and individual school influences such as policy structures, social climate, resource constraints, federal and state mandates, and state tenure laws.

The *inputs* are (a) the *district and school inputs* including evaluation policies, role definitions and assignments, evaluation budget, and evaluation timetable; and (b) the *enabling conditions* in the district whose presence assists the operation of the teacher evaluation system and whose absence likely impedes the teacher evaluation operations, e.g., management, school climate, training of evaluators, involvement of the teachers organization, and periodic review and improvement of the evaluation system. The actual teacher *evaluation process*, includes delineating teacher responsibilities; obtaining and documenting data and judging performance; providing formative feedback and reporting summative results; and applying information to guide professional development or to inform personnel decisions.

The *products* include (a) the *quality of evaluation results*, i.e., propriety, utility, feasibility, and accuracy; and (b) the *influences of the evaluations* (including uses, lack of use, and misuses of results) on individual professionals and groups of professionals, on the school district and individual schools, and on students and parents.

Figure 3-1 is a general overview of the variables found in the study of actual teacher evaluation systems. It is used here as the guiding conceptual framework for examining and improving teacher evaluation systems. Later a detailed form is

Figure 3-1. Conceptual Framework: Factors that Define and Influence Performance Evaluation Systems



presented reflecting the structure of Figure 3-1 for use in characterizing an existing teacher evaluation system. That form is a checklist designed to help school district teacher evaluation improvement teams to describe their current evaluation system in terms of Figure 3-1. By using the checklist, the team can identify ambiguities as well as the clear characteristics of their present system. When the team reaches agreement on what the current system actually is, it can proceed to diagnose its strengths and weaknesses. Figure 3-1 is a guide for examining and improving teacher evaluation systems. It helps respond to the next question.

9. How Can a District Use Figure 3-1 In Conjunction With The Personnel Evaluation Standards to Diagnose the Problems in its Teacher Evaluation System and Develop a Plan for Improving the System?

The scheme in Figure 3-1 provides a comprehensive perspective on the nature of teacher evaluation, including the things that might influence the way it is conducted and determine the quality of its findings and the extent of its influences. It is useful to view the scheme as a guide not only for describing an existing teacher evaluation system but also for diagnosing the system's strengths and weaknesses.

In the diagnostic process it is appropriate to focus first on the scheme's core category, the **Assessed Evaluation Service** which helps judge the **quality** of evaluation results or products. In addressing this core category one needs to determine which of the 21 Joint Committee standards are met or not met by the present teacher evaluation system. Judgments of the system against each standard are essential to decide whether the system meets the fundamental conditions of propriety, utility, feasibility, and accuracy. In a later section we provide detailed forms and advice on how to judge a teacher evaluation system against the requirements of each of the Joint Committee standards.

As an extension of the assessment of the teacher evaluation system's **propriety, utility, feasibility, and accuracy** and the extent to which each of the 21 personnel evaluation standards is met, one focuses next on the other impact category in Figure 3-1, which helps identify both desirable and undesirable influences of the district's teacher evaluations.

Teacher evaluations can influence individual teachers, groups of teachers, the institution or system in which the evaluation has been conducted, or its clients—parents, students, and the general public. In examining influences of teacher evaluation, one might ask the following questions: what are the positive and negative impacts on individual teachers, groups of teachers, the institution, the students and other customers? What negative consequences must be eliminated? What additional positive consequences should be sought?

Once it has been determined that the quality and consequences of the teacher evaluation system are deficient, then it is appropriate to consider why this is so and to identify corrective actions. As indicated above one can start the diagnostic/prescriptive process by identifying errors listed for each violated standard that are or seem to be a problem in the system. Similarly, one can identify the guidelines listed for the violated standards that if followed would help correct the noted deficiencies. These steps yield the initial working hypotheses about what to correct and how to proceed. They also underscore the importance of using the Joint Committee *Personnel Evaluation Standards* in conjunction with this GUIDE.

Next, one can use Figure 3-1 to extend the systematic search for the reasons that underlie the teacher evaluation system's poor outcomes. These reasons may be found by examining the context, input, and process variables in Figure 3-1.

The order of examining these variables is not critical so long as all are considered and the examiner is alert to possible contingency relationships among the categories of variables as well as individual variables. Thus, it is appropriate to consider Figure 3-1 as a learning guide rather than a mechanical device for systematic examination of a teacher evaluation system. For convenience the context, input, and process categories of variables are described below in their order of appearance in the figure.

The *state/community context* includes (among other variables) student characteristics and needs; state requirements for teacher evaluation, including the state teacher tenure and collective bargaining; laws; district policies, goals, and priorities; community attitudes toward educational excellence; social climate in the community; and the availability of funds for education. Obviously, many of the social context conditions are beyond the control of the school district. Since they can greatly influence and constrain a district's success in its teacher evaluation system, however, the context variables must not be ignored. In evaluating its teacher evaluation system, the district needs to note the context constraints under which they had to operate; also the areas of external support that should be used, perhaps better than the district now uses them; areas where the district might need to lobby for changes in state teacher evaluation policy; the expectations of parents and community; and, most important, the needs of its students.

District/school inputs include things in the school system that it can control and that are required for a workable, effective process of teacher evaluation to exist. These institutionally-controlled inputs greatly assist or constrain the teacher evaluation process. Included in this category are the presence or lack of many appropriate printed materials: a clear rationale for evaluation, evaluation policies, a general evaluation model, assigned responsibilities for participation in evaluation, defined duties and job descriptions for teachers, defensible measurement tools and techniques, and budgeted resources for implementing the evaluation.

If an evaluation process is not working or if the results in meeting any of *The Joint Committee Standards* are poor, then it is important to look at the adequacy of

all the above district inputs to see if basic changes are necessary and possible at that level. It won't do much good to work on the process if it can't possibly succeed under the present constraints in institutional inputs, such as unclear policy, ambiguous teaching assignments, and inadequate budget.

Enabling conditions are those dynamic inputs in the district whose presence or absence assists or thwarts the carrying out of an efficient teacher evaluation process but are not directly part of the process. Enabling conditions are the positive, supportive processes needed to make evaluation work. They include effective oversight and control of the evaluation process; a concept of teacher evaluation that is not only sound and in writing, but also known and endorsed throughout the district; a pervasive orientation to serve students and assist teachers; regular training of evaluators; evaluators who are not only well trained but also trusted and respected; a tradition of healthy interaction between supervisors and teachers; respect and support of the teacher evaluation system by parents, teachers, board members, and other stakeholders; regular access to sufficient funds and other resources to fully implement the teacher evaluation system; periodic review and improvement of the evaluation system; and constructive involvement of the teachers' organization. While the above conditions are not directly a part of the process of evaluating individual teachers, their presence or absence can substantially influence the effectiveness of the evaluation process and the extent and quality of evaluation outcomes.

The *evaluation process* is implemented through the major tasks of *delineating* the evaluation questions, intended uses, and required information; *obtaining* the information; *providing* the findings; and *applying* the results. These four tasks encompass a number of specific steps, as delineated in Figure 3-1. Among these are setting and maintaining a clear, feasible schedule for performance evaluation; clarifying intended users and uses of the evaluation; clarifying and validating teacher role definitions and performance criteria and standards; measuring and judging teacher performance; obtaining stakeholder input; considering information on the work environment and student needs; documenting the evaluation procedures; communicating evaluation results; controlling bias in the evaluation process; keeping appeal channels open; controlling the distribution, storage, and use of evaluation reports; and making decisions based on the results.

These steps reveal the complexity of the teacher evaluation process and identify a number of aspects where it can go wrong.

10. In General, What Overall Process Is Involved in Applying the Model in Figure 3-1 to Analyze a Teacher Evaluation System?

In examining a teacher evaluation system one looks first at the quality of results and the extent and desirability of influences on individuals and groups. Are the evaluation reports on individual teachers clearly grounded in sound information about the teacher's job performance and is the information effectively used to help teachers improve and to terminate the persistently incompetent teachers? If the evaluations of individual teachers are not functioning satisfactorily, one then might examine the context, input, and process categories to determine specific deficiencies and what variables can be affected in order to improve the teacher evaluation reports and their impact. Looked at this way, Figure 3-1 provides a framework for formulating working hypotheses about how to strengthen particular teacher evaluation systems. Later in this chapter detailed illustrations are presented for using Figure 3-1 in the process of diagnosing the strengths and weaknesses of a teacher evaluation system as a basis for improvement efforts.

Finally, it should be noted that an additional and significant benefit of the process described above is that it helps a team to develop a common view of such important evaluation components as teacher duties, standards for evaluation systems, purposes of the local system, and areas in need of strengthening.

The purpose of this section has been to suggest a state-of-the art conceptionalization of what teacher evaluation is and should be, and what steps are involved in improving a teacher evaluation system. In summary, the most important steps in improving a teacher evaluation system are as follows:

1. Adopt the Joint Committee Standards as a policy for assessing and improving the teacher evaluation system.
2. Staff the improvement effort with a representative body, and use a democratic process to apply the Standards and improve the evaluation system.
3. Carefully document the present evaluation system (focus, rationale, uses, policies, questions, performance criteria, procedures, materials, reports, timing, frequency, budget, etc.).
4. Apply the Joint Committee Standards to the evaluation system to identify which standards are being met and which ones are not.
5. Examine the common errors and guidelines in each unmet standard to identify and help diagnose the flaws in the present evaluation system.
6. Use the scheme in Figure 3-1 to extend the diagnosis.
7. Develop a plan and obtain support for implementing a project to improve the evaluation system.

The GUIDE now turns to presenting more operational advice for carrying out each step.

Organizing a Participatory Project to Improve Teacher Evaluation

It is crucial to involve all stakeholder groups in organizing the assessment of the adequacy of the school district teacher evaluation system. Nothing of significance is likely to be accepted, accomplished, or sustained in such an activity unless all those affected by it are kept informed and provided access to appropriate involvement. In addition, with so complex an issue, the results are likely to be more complete, relevant and useful if all possible know-how and perspectives are brought to bear.

The Joint Committee on Standards for Educational Evaluation put it well in the following statement:

If personnel evaluation policies and procedures are understandable, cooperatively developed, acceptable to all interested parties, and officially adopted, they are likely to assure continued cooperation within the personnel evaluation program. Such cooperation fosters support for the program, commitment to its purposes, acceptance of its methods, effective implementation, confidence in the reports, and trust in evaluation outcomes (Joint Committee, 1988, page 75).

The remainder of this section describes four specific steps in organizing a teacher evaluation improvement project adopting that perspective.

1. Get a "Go Ahead" from Key Agencies

Prior to solicitation of involvement of all stakeholder groups, the official structures within the system that make decisions on and determine resources for such activities must commit themselves.

Among these groups, the most obvious and prominent are the school district governing board, the district administration, and the local teachers' organization. In addition, if the school system is large enough to support a separate staff devoted particularly to personnel matters, this unit needs to be involved from the beginning.

From whatever source the idea originates for evaluating the district teacher evaluation system—teachers, administrators, school board, the community, or other—the initiating group needs immediately to respond to the following:

- What agencies within the school system must give official approval for such a project to get under way?
- What is the best estimate of financial resources required?
- What kinds of staffing will be required?
- How much staff time will be required?
- What time commitments will be required of those who compose the evaluation group?
- What duration of time will be required for such a project?
- Who will be the most important parties, both formal and nonformal, in getting the project under way and sustaining it throughout?

When the initiating group has answered the above questions, it should approach whatever decision-making person or group in the school district it considers most important for developing interest and commitment. A brief, to-the-point written position statement on what needs to be done, why, through what auspices, and what it will contribute to overall educational improvement may be useful for establishing interest and commitment. In cases where excellent rapport and communication already exist between the parties, it may be best to consider any written plan as tentative and keep it in the background while jointly outlining basic agreements for the proposed teacher evaluation improvement process.

In any case, the superintendent or other key decision maker will undoubtedly request that a more formal prospectus be prepared to confirm the project's foundation agreements. This prospectus should be designed to assist both the lead decision maker and the initiating groups to inform stakeholders about the nature, importance, and general outline of the project. Questions like the following should be answered in the prospectus:

- What group authorized the project, e.g., the board and superintendent?
- What needs and problems will this project address, e.g., invalidity and lack of credibility of the present teacher evaluation system?
- What special opportunities will be used, e.g., involvement in a pertinent research and development project on teacher evaluation?
- What are the project goals, e.g., systematic evaluation of the present teacher evaluation system; design of an improved system; testing, revision, and validation of the new system; installation of the new system?
- What group is providing conceptual and management leadership to the project, e.g., the district's research and evaluation department and a collaborating national research and development center?
- How will interested stakeholders have access to project information and involvement, e.g., regular briefings and a project advisory board?

- What is the time line and overall schedule of work for achieving the project goals?
- Who will do the work and how much of their time will be required?
- How will the project be monitored and evaluated, e.g., through an external metaevaluation panel?
- What is the budget and source of funds for the project?

When a decision to proceed has been made by the appropriate official body, the time will come to constitute an improvement team. Such a group might be used mainly as a sounding board and communication channel to all interested stakeholders, or it might have a more authoritative, active role. In the latter case, it would give overall leadership and guidance to the effort and conduct much of the work. The school district leadership should carefully consider what role is best for the improvement team, then clearly define its charge to the team. A list of all stakeholder groups in the school district should be assembled, and project leaders should fan out to explain and "sell" the improvement project. In doing so, individuals within each body, group, or institution should be identified who have interest in and also, as possible, background and experience that promise to enhance the effort.

In this way, a tentative roster of membership for an improvement team can be developed. Official appointment should be made by the school district governing board or superintendent. With a tentative representative group identified, activities can be planned and initiated for members of that group as well as any other work group to develop an understanding of *The Personnel Evaluation Standards*.

2. Initial Activities for Coming to Understand the Standards

At least five things need to be accomplished during a first convening of the project oversight and other work group(s) (following such organizational activities as electing officers and setting meeting times):

1. Provide an overview of the purpose of the project.
2. Provide key members of the teacher evaluation improvement team with a copy of this guide.
3. Present an overview of Sections Two (on the purpose, meaning, and improvement of teacher evaluation) and Three (on organizing a project to improve teacher evaluation) of the GUIDE.
4. Provide a copy of *The Personnel Evaluation Standards* to each team member.
5. Provide guidelines to team members on how to study the *Standards*.

Guidelines for studying the *Standards* should include the following steps:

1. Read the statement of the standard.
2. Study the rationale and guidelines for the standard (understanding the guidelines is particularly important, since they are one of two main sources of a basic information-gathering instrument to be used in accomplishing the purposes of the project).
3. Study intensively at least *one* illustrative case for each standard (similar to the illustration in Part 2 of *The Personnel Evaluation Standards*).

This last activity can bring "real world" understanding of what results when a particular standard is *met* or *not met*, since the illustrative cases describe what actually takes place in educational institutions in relation to meeting specific standards.

3. Practice in Applying the Standards

In teaching district personnel about the *Standards*, it can be useful to conduct a group exercise in which the participants apply the *Standards* cooperatively in assessing some illustrative evaluation model or system. They might, for example, read a description of some other teacher evaluation model (such as those authored by Edward Iwanicki [1990], Madeline Hunter [1988], or Tom McGreal, [1983]), or review materials from another school district's actual teacher evaluation system.

Each group member can be assigned to read, apply, and teach two or three of the individual standards as they apply to the particular evaluation model. Thus, different group members will simultaneously read different standards. In applying each assigned standard, the group member lists the evaluation model's strengths and weaknesses in relation to the standard's requirements and record a judgment of whether the standard is met, partially met, or unmet. Subsequently, each group member will teach the others the substance of each standard that he/she applied and report on its application to the teacher evaluation model under review.

Then the group will review their collective results from applying all 21 standards and through a consensus process make judgments about the evaluation model's propriety, utility, feasibility, and accuracy. In addition, they could agree on which standards should receive priority attention in steps to strengthen the evaluation model.

This analysis will provide a basis for deciding whether the evaluation model is adequate as it stands, needs to be improved, or is so deficient that it should be rejected totally. If the decision is to improve the model, the above analysis will provide working hypotheses about which standards most need to be addressed in order to make the evaluation model fully satisfactory.

4. Clinching the Understanding

Following their initial attempts to learn and apply the Standards, it is useful to convene the oversight and work groups to review and "clinch" the learning. This is the time for a group that will work together for a considerable period to achieve common understanding of the meaning of the *Standards*, from the broad implications of their use for improving the local evaluation system to precise definitions of specific terms. For accomplishing the latter, members might be referred to the glossary in the *Standards* book. The above description, incidentally, reflects how the authors of this GUIDE have taught real groups the substance and application of the *Standards*. The process works.

In summary, the main points of this section are as follows:

1. *The Personnel Evaluation Standards* are the most widely endorsed standards in the U.S. for judging teacher evaluation and other educator evaluation systems⁵.
2. Assessment and improvement of a teacher evaluation system should be a collaborative activity of the interested stakeholders.
3. The stakeholders can be greatly aided in their capacity to work together if they teach each other about *The Personnel Evaluation Standards* and if they jointly use the *Standards* to examine and strengthen the school district's teacher evaluation system.

The oversight and project work groups are now prepared to take up the task of describing their district's evaluation system. This activity is presented in the next section.

Profiling the Current Teacher Evaluation System

It is assumed at this point that the Teacher Evaluation Improvement Team has gained a high level of understanding of *The Personnel Evaluation Standards*, has reached some knowledge of the other main concepts presented to this point in the GUIDE, and has achieved credibility and support for leading the district's effort to assess and, as appropriate, improve the teacher evaluation system. In that context, the purpose of this section is to provide practical advice and materials for the team

⁵ In fact, the Joint Committee on Standards for Educational Evaluation is the only body in education that has been accredited by the American National Standards Institute (ANSI) to set professional standards for educational evaluations in the U.S.

to use in performing its first assessment task: *describing the current teacher evaluation system*.

The aim of this step is to develop a shared understanding among members of the team and other stakeholders of the local teacher evaluation system as it is defined and as it is actually practiced. This involves clarifying and validating the assumptions about what the system is, both on paper and in practice. With a clear description of the design and operation of the evaluation system gained from implementing this section, the team will be able to apply *The Personnel Evaluation Standards* to the local teacher evaluation system. It cannot be overemphasized, however, that having a clear, valid understanding of the system is absolutely essential to an effective application of *The Personnel Evaluation Standards*. Before the team can evaluate the teacher evaluation system, however unclear its policies and practices might be, the team members must determine what it is.

1. Securing Information on the System

The team will need to search for information that is often hard to find as it seeks to document and describe the current teacher evaluation system. In this endeavor, the team's make-up can be critical. It is expected, for example, that among its members there will be those with intimate knowledge of school board policy, school system rules and regulations, and other sources of information on teacher evaluation. They can contribute to obtaining pertinent print documentation, a crucial source of information, since it contains the official district position and examples of past practice, as contrasted to the oral history and folklore about the system. Print documentation (or the lack of it) also attests to whether or not the teacher evaluation system has been clarified in specific policies, rules, regulations, and practice guidelines.

In addition to collecting and examining relevant documents, it is also important to investigate and record what actually happens in the practice of evaluating teachers. This information can fill gaps not covered by documents, and it can help determine whether evaluations are conducted in accordance with the written evaluation policies and guidelines.

2. Print and Practice Sources

The two main sources of information on the teacher evaluation system, *print* and *practice*, are described below.

1. *Get it in writing if possible*

Written evidence (print), such as the collective bargaining contract, is an essential source for identifying the teacher evaluation policies. Documents, such as teacher evaluation instruments and reports, are also useful to assess whether the evaluation policies and practices are consistent.

Examples of written sources of information include the following (the team will likely identify others):

- written school board policy
- written rules and regulations (based on policy)
- negotiated agreements
- collective bargaining letters of agreement
- job descriptions
- role definitions
- central office administrative handbooks
- contracts or letters of employment
- statements of responsibilities
- faculty handbooks
- principals' handbooks
- special administrative orders
- periodic bulletins to staff
- procedures manuals
- reporting forms
- rating forms
- memos and directives
- meeting minutes

2. *Get additional information from interviews*

When it is not possible to answer a particular question from print sources, oral inquiry (interviews) should be made to establish what the unwritten policies are and the extent to which the written policies are or are not being implemented. The most valuable product of such interviews usually is descriptions of actual evaluation practices from the perspectives of key participants. Often that is a reality quite different from what is described in written policies.

Personnel who might be interviewed include those who have the major responsibility for implementing the evaluation system (the team will likely identify others):

- director of personnel
- other staff in the personnel office

- principals
- other school building personnel directly involved in the evaluation process
- other school staff who have some responsibility for implementing rules and regulations that bear on teacher evaluation

Teachers, the subjects of the evaluation, and principals, most often the evaluators, are particularly good sources of information about the actual practice in the district. Normally they can provide illustrative materials drawn from actual performance evaluations, and in addition personal descriptions of their experiences with the system. A form for use in recording findings from both print and practice is part of Appendix A referred to below. Print and practice also are essential background sources required for the completion of the forms in Appendix B.

3. Use the Collected Documents and Interviews to Indicate the Characteristics of the Evaluation System on a Common Form

Once the background information on the teacher evaluation system has been obtained, the team is advised to code the system on an appropriate form. Appendix A contains such a form developed from a study of a range of different teacher evaluation systems. This form is comprehensive in its coverage of variables found in those systems. It will enable the team to record what it has learned about the current teacher evaluation system related to the following major questions, as well as more specific issues:

- In what school district/subset of schools is the evaluation system employed and, which teachers are subject to the evaluation?
- Who developed the evaluation system and what is the nature and level of involvement of the teachers' organization?
- Are there official policies for teacher evaluation and what are they?
- What is the multiyear schedule for evaluating different classifications of teachers?
- What are the purposes of the evaluations?
- Who conducts and is otherwise involved in evaluations of teacher performance and what are their qualifications?
- What criteria are used to evaluate teacher qualifications and performance, and to what extent are work environment variables considered?
- What forms, instruments, procedures, etc. are used to evaluate teacher performance?
- What reports and other types of feedback are given to teachers, and how is the distribution of reports controlled?

- How are evaluation findings used?
- How, if at all, is the evaluation system monitored and modified?
- What, if any, groups of teachers are exempt from evaluation?
- What, if any, published model(s) provides the basis for the evaluation system?

Using the FORM FOR DOCUMENTING A TEACHER EVALUATION SYSTEM (Appendix A) will enable the teacher evaluation improvement team to gain a comprehensive view of the school district's teacher evaluation system. Involvement of the entire team and other stakeholders in completing the form will both enhance the comprehensiveness of documentation and promote shared understanding. Nevertheless, there probably will remain some gaps and areas of uncertainty.

Therefore, it is desirable that the team meet with other stakeholders, e.g., teachers, personnel office staff, principals, central office administrators, and board members, to present the completed form for further clarification and validation. A bonus of such activity is that involvement of a wide range of stakeholders in this process broadens the base of shared understanding and acceptance beyond membership of the team. This is an important benefit that can pay off through later widespread involvement in and approval of the process to improve the teacher evaluation system.

4. Abstracting the Teacher Evaluation System

Once the team is satisfied that it has collected sufficient documentation of the teacher evaluation system and correctly coded its characteristics on the FORM FOR DOCUMENTING A TEACHER EVALUATION SYSTEM (Appendix A), it should describe the system in narrative. The 13 questions listed above can serve as an outline for the description, or the team might create its own outline. If the team can develop and agree on a coherent, succinct description of the evaluation system, then its members can feel comfortable that they have achieved a shared understanding of the current teacher evaluation system. If not, they probably are not prepared to move ahead.

At this point the team is ready to advance to the next crucial step: Evaluating the current teacher evaluation system against *The Personnel Evaluation Standards*, the topic of the next section.

Determining which Personnel Evaluation Standards Are Met by the Present Evaluation System

When agreement on the description of the school district's evaluation system has been reached, *The Personnel Evaluation Standards* can be applied to assess the system, in order to determine the extent to which it possesses the desired attributes—propriety, utility, feasibility, and accuracy. This section presents two different approaches to that key process.

The first option to applying the *Standards* is the one recommended by the Joint Committee in Part II of the *Standards* book (Joint Committee, 1988, pages 123-154). This option is titled the Matrix Sampling Method. In it, the set of 21 standards is divided among the members of the team. For example, each member might be assigned 2 or 3 of the 21 standards. It is desirable to assign the *Standards* to insure that each standard is applied independently by at least 2 team members. This will provide cross-checks on the conclusions reached. Each member then systematically applies the assigned standards to the description of the district's current teacher evaluation system.

The steps in applying each standard are as follows:

1. Read the assigned standard definition, explanation, guidelines, and common errors.
2. In consideration of the requirements of the standard, make lists of strengths and weaknesses of the system. It is suggested that these lists be made on a form such as that provided in Figure 3-2.
3. Based on the noted strengths and weaknesses, record on the form a judgment of whether the standard is met, partially met, not met, or cannot be assessed.

If the judgment that the standard under consideration is either partially met or not met, the team member should list ideas for improvements that might strengthen the teacher evaluation system in meeting the standard. These lists of possible improvement steps will prove useful later when the team addresses the issues of redesigning the system.

If information is insufficient to apply the standard, the existing evaluation system can be considered to be incomplete. Until the needed information is obtained, it is appropriate to judge that the standard is not met.

When team members have applied all of the individual standards and recorded strengths, weaknesses, and recommendations for each, the team should convene and implement the following steps:

1. For each standard, those members who applied it should report their findings.

Figure 3-2. Individual Standard Summary

Standard: Letter-Number Standard Title: _____

	WEAKNESSES	IMPROVEMENT RECOMMENDATIONS

JUDGMENT CHECKLIST: The Standard is:

- ☐ Met
- ☐ Partially met
- ☐ Not met
- ☐ Not applicable
- ☐ Insufficient information

2. The full team should hear and discuss the reports for each standard; reconcile discrepancies; merge the different lists of strengths, weaknesses, and recommendations for the standard; and reach consensus on whether the standard is met, partially met, or not met.
3. A team recorder should document the team decisions for each of the 21 standards on forms, such as Figure 3-2.
4. The team should then summarize the overall profile of the teacher evaluation system relative to the 21 standards. See Figure 3-3 for an example of a summary form and Figure 3-4 as an example of a profile sheet that the recorder might use to record the overall team judgments of the system against the 21 standards.

A **second approach** recommended by the Joint Committee (1988) would be to have each team member independently apply all 21 standards. Using a separate individual form (Figure 3-2) for each standard, each member would list strengths, weaknesses, and recommendations for improvement and record her or his judgment of whether the standard was met, partially met, or not met. Each member would then record those judgments on a summary form.

This homework by each team member is designed to prepare her/him for informed participation in a group consensus process. After all members have completed the individual analysis of all standards, they would submit their forms to the team leader. The leader, with appropriate secretarial support, would summarize for each standard the information from individual team members. The summary (see Figure 3-3) would include a merged list of all noted strengths, weaknesses, and recommendations and would identify the number of team members who judged the standard to be met, partially met, or not met. Each team member would then be provided both her/his individual sheet and the summary sheet for each standard. The team would then meet to do the following:

1. Discuss each standard to reach agreement on whether it is met, partially met, or not met.
2. Decide whether any of the listed strengths, weaknesses, and recommendations should be deleted.

Either of these approaches should assist the teacher evaluation improvement team in determining the extent to which the teacher evaluation system meets *The Personnel Evaluation Standards*. This chapter and the *Standards* book should provide sufficient direction and information for accomplishing the task.

Once it is determined that the results of applying the *Standards* are valid, team members should discuss those results to identify the critical issues and objectives to be achieved in strengthening the evaluation system. Essentially, the team

Figure 3-3. Standards Summary

STANDARDS	STRENGTHS	WEAKNESSES	JUDGMENT*	RECOMMENDATIONS
P-1 Service Orientation				
P-2 Formal Evaluation Guidelines				
P-3 Conflict of Interest				
P-4 Access to Personnel Evaluations				
P-5 Interactions with Evaluatees				
U-1 Constructive Orientation				
U-2 Defined Uses				
U-3 Evaluator Credibility				
U-4 Functional Reporting				
U-5 Follow-Up and Impact				

* JUDGMENT CHECKLIST: The Standard is met, partially met, not met, not applicable, insufficient information

Figure 3-3. Standards Summary (continued)

STANDARDS	STRENGTHS	WEAKNESSES	JUDGMENT*	RECOMMENDATIONS
F-1 Practical Procedures				
F-2 Political Viability				
F-3 Fiscal Viability				
A-1 Defined Role				
A-2 Work Environment				
A-3 Documentation of Procedures				
A-4 Valid Measurement				
A-5 Reliable Measurement				
A-6 Systematic Data Control				
A-7 Bias Control				
A-8 Monitoring Evaluation Systems				

*JUDGMENT CHECKLIST: The Standard is met, partially met, not met, not applicable, insufficient information

Figure 3-4. Evaluation System Profile

	Addressed and Met	Addressed and Partially Met	Addressed and Not Met	Not Addressed
PROPRIETARY STANDARDS				
Service Orientation	☐	☐	☐	☐
Formal Evaluation Guidelines	☐	☐	☐	☐
Conflict of Interest	☐	☐	☐	☐
Access to Personnel Evaluation Reports	☐	☐	☐	☐
Interactions with Evaluatees	☐	☐	☐	☐
UTILITY STANDARDS				
Constructive Orientation	☐	☐	☐	☐
Defined Uses	☐	☐	☐	☐
Evaluator Credibility	☐	☐	☐	☐
Functional Reporting	☐	☐	☐	☐
Follow-Up and Impact	☐	☐	☐	☐
FEASIBILITY STANDARDS				
Practical Procedures	☐	☐	☐	☐
Political Viability	☐	☐	☐	☐
Fiscal Viability	☐	☐	☐	☐
ACCURACY STANDARDS				
Defined Role	☐	☐	☐	☐
Work Environment	☐	☐	☐	☐
Documentation of Procedures	☐	☐	☐	☐
Valid Measurement	☐	☐	☐	☐
Reliable Measurement	☐	☐	☐	☐
Systematic Data Control	☐	☐	☐	☐
Bias Control	☐	☐	☐	☐
Monitoring Evaluation Systems	☐	☐	☐	☐

determines which standards will receive priority attention in the effort to improve. Thus, they develop the "big picture" of what is to be accomplished.

The following section addresses the team's next important task: Deciding and planning how to improve the teacher evaluation system.

Deciding and Planning How to Improve the Teacher Evaluation System

Up to this point, the Teacher Evaluation Improvement Team has described the current system and evaluated it against *The Personnel Evaluation Standards*. The entire team has discussed and agreed upon the degree to which the system meets, partially meets, or does not meet each of the 21 standards. The team also will have compiled preliminary lists of strengths, weaknesses, and recommendations related to each standard. With these determinations, they will have reached their conclusions about the present system's propriety, utility, feasibility, and accuracy.

The team's next task is to decide how best to improve the teacher evaluation system. They have three basic options:

1. Decide that the present system is exemplary in satisfying the conditions of propriety, utility, feasibility, and accuracy. In this unlikely but most welcome case there is no need for corrective action. And, indeed, a celebration may be in order. Certainly, the district should make its outstanding evaluation system known outside its community, so that others may study and consider adopting the procedures that led to its success.
2. Decide that the present system is so seriously flawed in its propriety, utility, feasibility, and accuracy that it must be replaced with a new system.
3. Decide that the present system has sufficient merit to warrant its improvement and continued use.

Replacing the Current Teacher Evaluation System

In some cases it will be appropriate, even mandatory, that the team chooses the second option. Some teacher evaluation systems are so inappropriate or ineffective that they are beyond repair. For example, a number of states have instituted innovative teacher testing and career ladder programs, only to conclude later in the face of controversy, poor performance of the system, and litigation that the new system should be discontinued and replaced. Clearly, there is justification and precedent for drastic action.

However, making and implementing such a significant decision is complex and difficult. A host of issues cause difficulties in choosing to replace a teacher evaluation system:

- renegotiating the collective bargaining agreement on teacher evaluation
- dismissing or reassigning current staff
- discontinuing use of the current evaluation form, which may be favored by administrators for its convenience
- convincing stakeholders that there is genuine need for a change and that replacement will be professionally sound, not a political power move
- obtaining the funds, associated resources, and administrative commitment necessary to design, test, refine, and install a new system
- training participants to implement a new system
- convincing stakeholders that improved evaluation can and should serve many valuable ends other than perfunctory accountability uses

The district's teacher evaluation improvement team, central administration, and board will want to weigh carefully the pros and cons of the choice between replacing or improving the present system. The key determinant in that decision should be the assessment of which option is more likely to result in a system that meets the requirements of *The Personnel Evaluation Standards*.

If the board decides to discard and replace the current system, then it must set a timetable and budget for phasing out the present system while simultaneously designing, trying out, revising, and installing a new system. This will normally require one to two years. The necessary development and installation work should be carefully defined, assigned, scheduled, and budgeted, since this is a complex and sensitive process carried out in the shadow of a failed predecessor.

The new system should be designed with appropriate attention to all the steps in Figure 3-1. Here is a list of steps to be implemented in replacing the present evaluation system:

1. Gain go-ahead for replacing the teacher evaluation system from the official governing body and administration of the school system and at least a sign-off from the teachers' organization and other key formal bodies.
2. Identify all audiences that must become involved and/or fully informed as improvements are agreed on and implemented.
3. Ask the board to adopt *The Personnel Evaluation Standards* as district policy for evaluations of teachers and other district personnel.
4. Obtain an official written charge including the following:
 - a clear and precise mission statement
 - a description of the team's authority and specific responsibilities

- time lines for accomplishing agreed-on tasks
 - a statement of how, to what audiences, and at what intervals reports of progress will be made
5. Announce the decision that the present system is to be discontinued, specifying a target date one or two years hence.
 6. Announce that a new system will be developed and installed in accordance with the requirements of *The Personnel Evaluation Standards*.
 7. Name members of the teacher evaluation improvement team.
 8. Provide opportunities for stakeholders to give input to the teacher evaluation improvement process, e.g., schedule open forums and focus group meetings and extend invitations to submit written observations and suggestions.
 9. Project the kind and numbers of human and material resources required to design and install the new system.
 10. Estimate the time required to plan, test, and fully install the new system.
 11. Develop and publicize a schedule of the development process, including
 - identification and assessment of alternative teacher evaluation models
 - outlining of the model to be employed in the district
 - operationalizing the model in such terms as policy statements, rules, annual cycle, criteria, performance information, forms, reports, storage and retrieval of reports, follow-up actions, responsibilities, training, and facilities
 - develop and obtain approval for an appropriate budget
 - review and testing of the operationalized model
 - training and installation
 12. Make a detailed checklist of the conditions to be met by the new system, drawing from the strengths, weaknesses, and recommendations listed for each standard and the items included in Figure 3-1.
 13. Identify alternative evaluation models or systems operating in school districts and evaluate them against *The Personnel Evaluation Standards*.
 14. Select a model or combination of models.
 15. Develop a general description of the intended new teacher evaluation system and share it with stakeholders.
 16. Develop a specific plan for developing and installing the new system.
 17. Affix the following to the plan for each component:
 - Financial resources required for its development
 - Human resources required for its development
 - Time lines required, by stages, for its development
 - Evaluation criteria and processes to be used in assessing the success of the new system

18. Compile the specific plans for development into an overall plan and specific work schedule for establishing the new teacher evaluation system.
19. Develop the design and component parts of the new teacher evaluation system, reflecting the process steps and related components of Figure 3-1. What must be done in specifics will depend much on the particular model selected or constructed and on local community and school district conditions. For illustrative purposes, some of the techniques that cut across a variety of teacher evaluation models are listed below for the consideration of implementers. They are placed generally in the order in which they might be considered if an evaluation system were being developed from the beginning:
 - Validate the need for the existing teaching positions and make recommendations for appropriate changes.
 - Update teachers' job descriptions.
 - Specify general performance standards and criteria.
 - Define evaluation report formats.
 - Select or develop evaluation forms and other materials, as needed.
 - Define evaluation uses and users.
 - Define roles and responsibilities for implementing the new evaluation system.
 - Provide needed orientation and training.
 - Schedule evaluation tasks.
20. Evaluate the operationalized new evaluation system design against the 21 *Personnel Evaluation Standards*. Follow the procedures that the team used to evaluate the present teacher evaluation system.
21. Revise the evaluation system design as appropriate.
22. Train the participants to implement the teacher evaluation system.
23. Formally install the improved system.
24. Periodically review and evaluate the system and improve it as appropriate.

A Word About Alternative Teacher Evaluation Models

Please note that the reviews of teacher evaluation models (Scriven, Wheeler, & Haertel, 1992-93; Stufflebeam, 1992; Dwyer & Stufflebeam, 1994) were not encouraging. The reviewers found that virtually all existing published teacher evaluation models are inadequate in comparison to the requirements of the full range of *The Personnel Evaluation Standards*. While many models do evidence decided strengths, virtually all of them also have serious weaknesses. Some models (e.g., the value-added models of Sanders & Horn, 1994; and Webster, Mendro, & Almaguer, 1994) are strong on technical grounds and on their use of student

performance data to assess teaching effectiveness, but lack feasibility in the great majority of school districts. Other models are easier to use in the wide range of districts, e.g., Management By Objectives and teacher self-assessment, but these often lack rigor and credibility in their application. Scriven's approach, which focuses on assessing a teacher's fulfillment of duties, is conceptually and philosophically compelling, but has not been operationalized as a component of an existing, or new, model. The frequent practice of annual or semiannual observations of teachers by principals often fails to discriminate between good and poor teaching throughout the year, focuses inappropriately on teaching style rather than responsibilities, and may not provide useful feedback to the teacher. Some models, such as Iwanicki's Professional Growth Oriented Model (Iwanicki, 1992), place so much emphasis on teacher improvement that they lack credibility with respect to identifying serious teaching deficiencies and dismissing persistently incompetent teachers.

In spite of the mixed report on the merits of alternative teacher evaluation models, the teacher evaluation improvement team can gain much of value by searching out and studying these models in comparison to *The Personnel Evaluation Standards*. The team should undertake this task with the aim of selecting strong features from several different models and combining them into a strong hybrid. Those sources marked with an asterisk (*) under *References* should be useful for identifying and examining alternative teacher evaluation models.

Improving the Current Teacher Evaluation System

In many cases, a district will determine that its present teacher evaluation system is sufficiently sound in propriety, utility, feasibility, and accuracy and that it is best to improve rather than replace it. In such instances, the teacher evaluation improvement team should ground its redesign in the results of its assessment of the current system against the 21 Joint Committee Standards.

Weaknesses identified when the evaluation system is compared to the *Standards* as stated on the individual standard form (Figure 3-2) provide a good starting place for determining needed corrections. The strengths found in the current system should also be considered and built upon. And it will be useful to examine the recommendations that were recorded. The previous analysis of the present system in comparison to the 21 *Standards* provides an important list of specific, germane issues to be resolved and strengths to be built upon in planning the improvement of the evaluation system.

In addition, the structure in Figure 3-1 offers a framework by which to organize the identified strengths, weaknesses, and recommendations related to the components of a teacher evaluation system. This figure essentially summarizes context,

input, process, and outcome categories and shows general interrelationships among the full set of identified variables for defining a teacher evaluation system and determining its strengths and weaknesses.

This section illustrates how Figure 3-1 can be used to formulate propositions about how to improve particular teacher evaluation systems. The propositions are presented separately for each of the four main principles in the quality paradigm (propriety, utility, feasibility, and accuracy) and are examined in relation to five other dimensions in Figure 3-1 (State/Community Context, District/School Printed Inputs, Enabling Conditions, Evaluation Process, and Impacts of the Evaluation). It is recommended that teams carefully develop their own lists.

The sample proposition (propriety) presented below is not an exhaustive catalog of ways to improve a particular teacher evaluation system. Rather, the intent is to provide an illustrative listing of potential deficiencies that should be addressed, as they are capable of causing the teacher evaluation system to fail. In using this scheme in actual teacher evaluation improvement projects, the team will develop its own propositions based on its comparison of the system against each Joint Committee standard and on how the identified strengths, weaknesses, and recommendations fit into the scheme's context, input, process, and output categories. The following is provided as an illustration of how the framework in Figure 3-1 can be used to identify and refine corrections to be made in improving an existing teacher evaluation system.

SAMPLE PROPOSITION (*Propriety*):

If the evaluation of teacher performance essentially **ACQUIESCES IN THE FACE OF SOME STUDENTS BEING SUBJECTED TO POOR TEACHING** (a clear violation of the Service Orientation standard), then it is especially important to check for and correct or counteract deficiencies in at least the following:

(STATE/COMMUNITY CONTEXT)

- 1.11 Ensure that everything that can be done has been done to secure the community's support and respect for competent teaching for all students.

(DISTRICT/SCHOOL PRINTED INPUTS)

- 1.21 Ensure that the rationale for teacher evaluation stated in the district policies places a high priority on assuring that every student will receive competent instruction.
- 1.22 Ensure that the district's printed policies are clear and defensible with respect to due process, remediation, and termination.

- 1.23 Ensure that the district policies require each teacher to be evaluated by a properly credentialed and trained evaluator.
- 1.24 Ensure that the stated purposes for evaluation include protecting students from substandard teaching.
- 1.25 Ensure that each teacher's duties are clearly defined in up-to-date official job descriptions.
- 1.26 Ensure that the sanctioned uses of evaluation include both remediation and termination.

(ENABLING CONDITIONS IN THE DISTRICT/SCHOOL)

- 1.31 Make sure the printed materials (e.g., manuals, memos, descriptions) on the evaluation system explicitly demand competent teaching for every student.
- 1.32 Make sure the printed evaluation system materials (beyond district policies), explicitly and consistently provide that teachers receive evaluative feedback oriented to improving teaching performance.
- 1.33 Make sure that the district has an explicit multiyear timetable (consistent with board policy) to assure that each teacher's performance is evaluated on a regular basis.
- 1.34 Make sure that reasonable efforts are made on a continuous basis to engage the teachers' union in supporting a teacher evaluation system that is strongly oriented to serving students.

(EVALUATION PROCESS IN THE SCHOOL)

- 1.41 Make sure that (a) the actual frequency of evaluations is sufficient to identify instances of substandard teaching before it harms students and (b) off-schedule evaluations are pursued when deficient teaching is suspected.
- 1.42 Make sure that work environment variables are examined for each teacher, so that steps can be taken to correct environmental problems and constraints that prevent a given teacher from succeeding.
- 1.43 Make sure that teachers are regularly given feedback on both teaching strengths and weaknesses.

- 1.44 Make sure that the district regularly employs due process for remediation and, if necessary, termination of teachers.

(IMPACTS OF THE EVALUATION)

- 1.51 Ensure that teachers are regularly given substantive feedback on the quality of their teaching along with suggestions for improvement.
- 1.52 Ensure that competent and especially exemplary teaching is recognized and reinforced.
- 1.53 Ensure that teacher evaluations lead to appropriate supervisory oversight and appropriate feedback.
- 1.54 Ensure that evaluations are used to help plan professional development activities.
- 1.55 Ensure that the evaluation is oriented to provide continuous improvement in service to students.
- 1.56 Ensure that the evaluation is oriented to ongoing surveillance of quality of teaching and equity of service.

The preceding sample proposition and identification of potential corrective actions shows how the framework in Figure 3-1 can be used in conjunction with *The Personnel Evaluation Standards* to develop specifications for improving an existing teacher evaluation system. The teacher evaluation improvement team is advised to measure its analysis of the system against the *Standards* and the contents of Figure 3-1 in order to develop a checklist of items that must be satisfactorily addressed to improve the system. They should find such a checklist useful both for redesigning the system and for evaluating the resulting plan.

The steps to be implemented in improving an existing teacher evaluation system are very similar to those listed earlier in this chapter for replacing a system, since the same factors and principles come into play. However, there are some important differences. Basically, the following steps are involved:

1. Gain go-ahead for improving the teacher evaluation system from the official governing body and administration of the school system, and at least a sign-off from the teachers organization and other key formal bodies.
2. Identify all audiences that must become involved and/or fully informed as improvements are agreed on and implemented.
3. Ask the board to adopt *The Personnel Evaluation Standards* as district policy for evaluations of teachers and other district personnel.

4. Obtain an official written charge including the following:
 - a clear and precise mission statement
 - a description of the team's authority and specific responsibilities
 - time lines for accomplishing agreed-on tasks
 - a statement of how, to what audiences, and at what intervals reports of progress will be made
5. Announce the decision that the present system is to be improved by a certain target date, one or two years hence.
6. Announce that the improvements will be made in accordance with the requirements of *The Personnel Evaluation Standards*.
7. Name the members of the teacher evaluation improvement team.
8. Provide opportunities for stakeholders to give input to the teacher evaluation improvement process, e.g., schedule open forums and focus group meetings, and extend invitations to submit written observations and suggestions.
9. Project required financial resources for work of the task force and accomplishment of improvements.
10. Project the kind and numbers of human and material resources required for making the needed improvements.
11. Estimate the time required to plan, test, and fully implement the improvements.
12. Publish a schedule of the development process, including
 - developing a checklist of requirements for the improved evaluation system
 - outlining the improved teacher evaluation system
 - operationalizing the improved system in such terms as policy statements, rules, annual cycle, criteria, performance information, forms, reports, storage and retrieval of reports, follow-up actions, responsibilities, training, facilities, and budget
 - reviewing and testing the improved system
 - training and installation
13. Make a detailed checklist of the conditions to be met by the improved system, drawing from the strengths, weaknesses, and recommendations listed for each standard.
14. Prioritize required improvements into such categories as
 - things that require immediate "fixing"
 - important needed improvements that should become second-level priorities
 - third-level improvements, the implementation of which can be spread over two, three, or more years
 - things that need fixing but that can become part of a long-range plan

15. Develop a specific plan for bringing about each required improvement, beginning with those classified under the "a" category above and proceeding through the hierarchy of priorities.
16. Affix to the plan for each improvement the following:
 - financial resources required for its accomplishment
 - human resources required for its accomplishment
 - time lines required, by stages, for its accomplishment
 - evaluation criteria and processes to be used in determining the success of the improvement
17. Compile the specific plans into an overall plan and work schedule for improving the teacher evaluation system.
18. Develop the design and component parts of the improved teacher evaluation system reflecting the process steps and related components of Figure 3-1. What must be done in particular will depend much on the specific needed improvements and on local community and school district conditions. For illustrative purposes, some of the techniques that cut across a variety of teacher evaluation improvement initiatives are listed below for the consideration of implementers. They are placed generally in the order in which they might be considered if an evaluation system were being developed from the beginning:
 - Validate the need for the existing teaching positions and make recommendations for appropriate changes.
 - Update teachers' job descriptions.
 - Specify general performance standards and criteria.
 - Define evaluation report formats.
 - Revise or replace evaluation forms and other materials, as needed.
 - Define evaluation uses and users.
 - Define roles and responsibilities for implementing the revised evaluation system.
 - Provide needed orientation and training.
 - Schedule evaluation tasks.
19. Evaluate the operationalized improved evaluation system design against the 21 Personnel Evaluation Standards. Follow the procedures that the team used to evaluate the present teacher evaluation system.
20. Revise the evaluation system design as appropriate.
21. Train the participants to implement the improved teacher evaluation system.
22. Install the improved system.
23. Periodically review and evaluate the system and improve it as appropriate.

Summary

The purpose of this section has been to provide practical advice for planning the improvement of teacher evaluation in a district. It was noted that the initial decision is to determine whether the present system needs to be *replaced* or *strengthened*.

If improvement is indicated, then the next decision is to decide whether to import a new model or strengthen the present system. If a new model is to be installed, a search should be made to identify promising models, and they should be evaluated against the 21 Joint Committee *Standards*. Based on the results of that evaluation, a model should be selected, or the best features of several choices combined into a hybrid model.

If the decision is made to improve the current model, the team should construct a checklist of the conditions that must be met in the improved system. It was noted that this checklist should reflect the strengths, weaknesses, and recommendations identified when the current system was evaluated against the 21 Joint Committee *Standards*. In addition, these should be organized and fleshed out by using the framework of Figure 3-1.

Once the new model or specifications for improving the present system are identified, the improvement process should be carefully outlined and implemented. Finally, sample steps for such a work schedule were provided.

Planning and Implementing the Evaluation Improvement Project

The activities proposed in the previous sections consist of organizing for, planning, and initiating action to replace or improve the school district teacher evaluation system based on weaknesses and strengths identified as a result of applying the *Standards* to the evaluation system. The purpose of this concluding section is to recap the main message of the *GUIDE* and to present general guidance for implementing the change process.

The section addresses the following topics:

- Guiding Principles for Improving Teacher Evaluation Practices
- How to Organize the Improvement Effort
- The Importance of Pilot Testing and Improvement
- Budget Considerations
- Ongoing Oversight

Guiding Principles

Whatever group (improvement team or a similar body) takes on the task of planning for improvement of the school district teacher evaluation system, several principles will need to be kept in mind:

1. Change in social agencies (schools are no exception) is often difficult to agree on and slow to accomplish.
2. Planning and implementing improvements need to involve representatives of *all* stakeholder groups.
3. Establishing priorities needs to be an early and important activity and must provide for
 - getting the most-needed things done first
 - accomplishing less urgent improvements in later stages
 - setting realistic time lines for various levels of priority
4. Provision needs to be made for resources, both financial and human, sufficient to accomplish the agreed-on needed improvements.
5. Provision needs to be made for personnel training and retraining required to effect and sustain change.
6. Provision needs to be made for piloting (field testing) whatever is to be done before broad implementation.
7. Provision needs to be made for addressing and correcting problems of implementation as they arise.
8. Assurance needs to be provided that implementations meet sound professional standards and legal requirements.

How to Organize the Improvement Effort

Bringing stakeholders into the planning process *early on* and making them active contributors to planning and effecting change may make the difference between success or failure. The worn cliché that people are unlikely to become enthusiastic about changes they are not part of or consulted about certainly applies here.

The improvement team must be confident of support for the teacher evaluation system improvement effort. And it is imperative that they secure a mandate for the work to be done and a commitment to an appropriate level of financial and other resource support to see the improvement process through to completion.

The concept of mandate can have at least two meanings. On the one hand, it can mean an authoritative order or command, especially a written one. For example, a mandate in a school system can be a decision made by the school board to review

and perhaps revise the teacher evaluation system. This is a "top-down" definition of mandate. However, a mandate can also refer to the desires of constituents expressed to a representative body as a directive for change. For example, a teacher or group of teachers within a school system may express, either through a teachers organization or through organizational channels, a desire to have the teacher evaluation system reviewed and perhaps revised. This is the "bottom-up" meaning of mandate.

Securing a mandate to improve the existing teacher evaluation system, as is used in this context, means achieving an agreement between the "top-down" administrative order and the "bottom-up" directive for change. The best assurance of success in such a sensitive effort is commitment to the teacher evaluation system improvement process by at least three key groups: the administrative body; the governing policy board of a school system; and the teachers. Moreover, although teachers and administrators are the major players in the effort, a genuine mandate must involve all stakeholders, for without representation and support from a broad base of stakeholders, effective change will be hard to come by.

Securing this mandate at the replacement or improvement stage of the process is even more critical than the mandate for reviewing the evaluation system. Replacement or improvement of the system involves change, and change can be, or seem to be, a serious threat to the existing structures and those most directly affected by the changes. Those who are a part of designing and implementing the improvement process are less likely to feel threatened and more likely to be seriously supportive and committed to change.

The improvement team, whatever its composition, will need to possess among its members those who have a thorough understanding of the school system, the key governance and decision-making structures of the community, and the community characteristics and inclination. Participation in the process by those who have that knowledge is bound to produce in the range of stakeholders a heightened sense of the personal, community, and professional benefits of improved teacher evaluation.

Early on, the improvement team will need to identify expertise among its members or among other groups and individuals willing to serve in a consulting capacity. And it will need to identify key change agents in the school system—existing structures and formal and informal leaders who play important roles whenever change is considered in the school district. Inclusion, not exclusion, strengthens support and the mandate for change.

The Importance of Pilot Testing and Improvement

The new or improved teacher evaluation system should be used only on a limited basis prior to full implementation. Allowing a unit of teachers, a single building, or an administrative unit to test the system provides an opportunity to see it in action, not just on paper. Concerns and problems that arise from the pilot test should be reported and recorded, so that at the end of the trial period specific changes can be made to improve the operation of the system.

Soliciting periodic feedback about the process during the pilot testing phase should help to ensure cooperation and increase trust, important factors when changes are being made. Providing opportunities for evaluators, evaluatees, and others to discuss operational problems emerging during the pilot testing phase has the added advantage of bringing to bear a broader, and more practical perspective on the changes that will be made.

Budget Considerations

The improvement plan budget must include the costs of personnel, services, and materials. If implementation of the plan requires the time of several support staff, substitute teachers and attendant release-time arrangements so that district teachers can participate, or stipends for stakeholders, these costs must be included in the budget. Services also may include fees, and associated costs of consultants (transportation, follow-up activities, training of district personnel). Additional materials may be required. The purchase of notebooks, handbooks, videotape, training guides, and other references may be necessary to carry out the improvement plan. If these expenses are included in the plan, the associated costs must be part of the budget.

Ongoing Oversight

Once the new or improved teacher evaluation system has been installed, the district should consider designating the teacher evaluation improvement team as a "guiding review panel." This panel would periodically review the operation of the new system and make recommendations for correcting problems and more generally for strengthening the system. Based on the considerable knowledge and understanding the task force will have accumulated, it could provide especially insightful feedback to the board, superintendent, personnel office, principals, and teachers. Its metaevaluation (i.e., evaluation of the evaluation system) work should be grounded

in *The Personnel Evaluation Standards*, this GUIDE, and the district evaluation policies. The process outlined in the section that addresses which professional evaluation standards are met by the present evaluation system is especially pertinent to the metaevaluation task.

From time to time it is appropriate and also useful to obtain independent metaevaluation from an outside evaluation expert. Whoever is given the oversight assignment, some group must carry out this function to sustain and improve the system put in place after such considerable effort.

Conclusion

Teacher evaluation should be one of the most important processes in any school district, since it can greatly influence the quality of instruction provided to students. Unfortunately, however, evaluation systems are often poorly conceived and/or implemented, and a vital opportunity to enhance quality is squandered. School districts need to take concerted action to assure that their teacher evaluation practices are as good as the state of the art allows. Accordingly, all districts should adopt and meet the requirements of the established professional standards for sound personnel evaluation in education.

References

- *Dwyer, C., & Stufflebeam, D. (1994). Evaluation for effective teaching. *Handbook of Educational Psychology*.
- *Gullickson, A., & Airasian, P. (1993). A model of teacher self-assessment. Paper presented at the annual meeting of the National Council on Measurement in Education, Atlanta.
- *Haertel, E. H. (1991). New forms of teacher assessment. *Review of Research in Education*, 17, 3-30.
- Hunter, M. (1988). Create rather than await your fate in teacher evaluation. In S. Stanley & W.J. Popham (Eds.), *Teacher evaluation: Six prescriptions for success*, 32-54. Association for Supervision and Curriculum Development.
- Iwanicki, E. (1990). Teacher evaluation for school improvement. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation*, 158-174. Newbury Park, CA: Sage.
- Iwanicki, E. F. (1992). *A handbook for teacher evaluation and professional growth in more productive schools*. Storrs, CT: The Connecticut Institute for Personnel Evaluation.

- Joint Committee on Standards for Educational Evaluation. (1988). *The personnel evaluation standards*. Newbury Park, CA: Sage.
- McGreal, T. L., (1983). *Successful teacher evaluation*. Alexandria, VA: Association for Supervision and Curriculum Development.
- *Millman, J., & Darling-Hammond, L. (1990). *The new handbook of teacher evaluation*. Newbury Park, CA: Sage.
- Reineke, R. A., Willeke, M. J., Walsh, L. H., & Sawin, C. R. (1988). Review of personnel evaluation systems: A local application of the standards. *Journal of Personnel Evaluation in Education*, 1, 373-378.
- Sanders, W. L., & Horn, S. P. (1994). The Tennessee value added assessment system (TVAAS): Mixed model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, 8(3).
- Scriven, M., Wheeler P., & Haertel, G. (1992, 1993). *TEMP A Memos*. Kalamazoo, MI: Western Michigan University, The Evaluation Center.
- *Scriven, M. (1990). Can research-based teacher evaluation be saved? *Journal of Personnel Evaluation in Education*, 4(1), 19-32.
- Scriven, M. (1988). Duties-based teacher evaluation. *Journal of Personnel Evaluation in Education*, 1, 9-23.
- *Scriven, M. (1994). The duties of the teacher. *Journal of Personnel Evaluation in Education*, 8(2).
- *Stufflebeam, D. L. (1992). Competing rationales and associated models and approaches for evaluating the performance of teachers. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.
- *Stufflebeam, D. L., & Nevo, D. (1994). Educational personnel evaluation. In T. Husen & T. N. Postlethwaite (Eds.) *The International Encyclopedia of Education* (2nd ed.). Oxford: Pergamon Press.
- Webster, W., Mendro, R. L., & Almaguer, T. O. (1994). Effectiveness indices: The major component of an equitable accountability system. *Studies in Educational Evaluation*.
- *Wise, A. E., Darling-Hammond, L., McLaughlin, M. W., & Bernstein, H. T. (1984). *Case studies for teacher evaluation: A study of effective practices*. Santa Monica, CA: Rand.

Form for Documenting a Teacher Evaluation System

Document Inventory

The purpose of the Document Inventory section is to provide a record of the teacher evaluation materials found in the district. Once completed, a copy of this part of the form should be attached to materials and documents used to complete this inventory.

On the list below, check off all materials and documents found for the school district/system. Make a note of any unusual conditions found in the file.

- ☐ the school's or district's collective bargaining agreement (if one exists)
- ☐ the school or district board policies on teacher evaluation
- ☐ defined teacher duties
- ☐ documents describing the teacher evaluation system
- ☐ examples of individual teacher contracts
- ☐ examples of teacher job descriptions
- ☐ past written reviews or references to published information on the teacher evaluation system
- ☐ relevant evaluation instruments and forms
- ☐ district/school building handbooks
- ☐ other, please identify _____

1. Evaluation System Identification

1.1 School district/system name: _____

School district/system location: _____

1.2 Name/label of the *teacher evaluation system* to be reviewed: _____

Name(s) of person(s) completing the inventory: _____

Date of inventory completion: _____

1.3 *Type of school or district covered by the teacher evaluation system (check all that apply):*

- ☐ Private
- ☐ Public
- ☐ Primary
- ☐ Upper elementary
- ☐ Elementary
- ☐ Middle
- ☐ Jr. high
- ☐ High school
- ☐ Secondary
- ☐ Unspecified

1.4 *Grade levels (between kindergarten and grade 12) covered by the teacher evaluation system:*

K 1 2 3 4 5 6 7 8 9 10 11 12

1.5 *Number of teachers covered by the teacher evaluation system:* _____

1.6 *Teachers covered:*

- ☐ Probationary teachers
- ☐ Tenured teachers
- ☐ Substitute teachers
- ☐ Classroom aides
- ☐ Itinerant teachers
- ☐ Other, please specify

2. Developers of the Evaluation System

2.1 What **groups** participated in developing the evaluation system (*check all that apply*)?

- ☐ Teachers
- ☐ Teachers organization
- ☐ District administrators
- ☐ School principals
- ☐ External consultants
- ☐ State education department
- ☐ Parents
- ☐ School board members
- ☐ Other, please specify _____

2.2 What is the involvement of the **teachers organization** with the evaluation system (*check all that apply*)?

- ☐ None
- ☐ Collective bargaining agreement covers teacher evaluation
- ☐ Evaluation criteria are negotiated with the union
- ☐ Evaluation methods are negotiated with the union
- ☐ Evaluation instruments are negotiated with the union
- ☐ Union represents teachers in grievances about evaluation
- ☐ Unspecified
- ☐ Other, please specify _____

3. Key Policy Provisions

3.1 Which of the following characterize the written **policies** that cover the teacher evaluation system (*check all that apply*)?

- ☐ No particular written policy is evident
- ☐ Covered by written school **building-level** policy
- ☐ Covered by written school **district** policy
- ☐ Covered by written **state** policy
- ☐ Other, please specify _____

3.2 Which of the following are addressed/specified/defined in the written policies and/or rules and regulations that govern the teacher evaluation system (*check all that apply*)?

- ☐ Exclusions of special categories of teachers (*specify*) _____
- ☐ Special provisions for probationary teachers
- ☐ Special provisions for substitute teachers
- ☐ Special provisions for itinerant teachers
- ☐ Different provisions for elementary and secondary school teachers
- ☐ Explicit teacher responsibilities/duties
- ☐ Frequency of required evaluations
- ☐ Limitations on distributing evaluation reports
- ☐ Required schedule for the evaluation steps
- ☐ Rules for storing and controlling access to evaluation information
- ☐ Clarification of *who* may access which evaluation reports
- ☐ The bases and procedures for removing evaluation information from the school or central files
- ☐ Explicit written safeguards for protecting the privacy of evaluatees
- ☐ Process for appealing a teacher evaluation
- ☐ Provision for submitting a written response that becomes part of the teacher's permanent file
- ☐ Required use of a board-approved evaluation form
- ☐ Requirement to identify and address conflicts of interest in individual teacher evaluations
- ☐ Requirement and provision for training evaluators
- ☐ Requirement that each teacher have an up-to-date job description
- ☐ Requirement that deficiencies requiring immediate attention be handled promptly and not postponed until the written evaluation
- ☐ Requirement that teacher performance be assessed in the light of assessments of available resources, working conditions, incentives, community expectations, and other context variables
- ☐ Requirement that evaluation system be periodically reviewed
- ☐ Other, please specify _____

4. Schedule for Evaluations

- 4.1 What is the usual ***schedule*** for performance evaluations for each of the following groups (*please briefly describe each schedule*)?

Probationary teachers: _____

Tenured teachers: _____

Substitute teachers: _____

Other, please specify: _____

5. Purposes of the Evaluations

- 5.1 Which are the stated ***purposes*** of the teacher evaluation system (*check all that apply*)?

- ☐ Motivate teachers
- ☐ Encourage and assist professional growth
- ☐ Provide feedback on strengths and weaknesses of performance
- ☐ Remediate deficient teacher performance
- ☐ Recognize excellent teaching
- ☐ Reward meritorious teaching (merit pay)
- ☐ Document and reward extra service (incentive pay)

- ☐ Assist the teaching profession to police and enhance its ranks
- ☐ Understand personal rôle in the school
- ☐ Monitor teacher performance in order to control and coordinate teaching across classrooms
- ☐ Inform personnel decisions (promotion, tenure, merit pay, termination)
- ☐ Develop competent teachers
- ☐ Maintain teacher accountability
- ☐ Safeguard student and community interests from incompetent or harmful teaching
- ☐ Assure high quality professional service to students
- ☐ Enhance student learning
- ☐ Enhance school credibility
- ☐ Unspecified
- ☐ Other, please specify _____

5.2 Which of the following **employment decisions** are served by the teacher evaluation system (*check all that apply*)?

- ☐ Selection of interns or student teachers
- ☐ Selection of new teachers
- ☐ Selection of support personnel
- ☐ Teaching job assignment
- ☐ Specification of job responsibilities
- ☐ Licensing/certification
- ☐ Confirmation of knowledge about the profession of teaching
- ☐ Confirmation of the teacher's basic literacy and numeracy skills
- ☐ Confirmation of proficiency with instructional techniques/methods
- ☐ Confirmation of proficiency with computer technology
- ☐ Confirmation of classroom teaching competence
- ☐ Confirmation of subject matter knowledge
- ☐ Continuation
- ☐ Issuance of notice to remedy
- ☐ Remediation
- ☐ Planning staff training and development programs

- ☐ Assignments to obtain special training or other individual staff development assistance
- ☐ Awarding of study leaves and special grants
- ☐ Promotion
- ☐ Tenure
- ☐ Special recognition
- ☐ Merit pay
- ☐ Incentive financial awards
- ☐ Rulings on grievances
- ☐ Sanctions
- ☐ Termination for cause
- ☐ Reduction in force
- ☐ Reorganization of teaching
- ☐ Unspecified
- ☐ Other, please specify _____

6. Responsibilities for Conducting the Evaluation

6.1 Who is involved in evaluating teacher performance (*check all that apply*)?

- ☐ School principal
- ☐ Head of department within school
- ☐ Committee of teachers from the school/district
- ☐ Self-evaluation by the teacher
- ☐ Team of administrators from the district
- ☐ District administrator or evaluator from outside the school
- ☐ Teachers from other districts
- ☐ Master teacher
- ☐ Groups of teachers from the teacher's school
- ☐ State inspector or evaluator
- ☐ School board
- ☐ Students
- ☐ Parents
- ☐ Unspecified
- ☐ Other, please specify _____

6.2 Who has the most important role in evaluating teacher performance
(check all that apply)?

- ☐ School principal
- ☐ Head of department within school
- ☐ Committee of teachers from the school/district
- ☐ Self-evaluation by the teacher
- ☐ Team of administrators from the district
- ☐ District administrator or evaluator from outside the school
- ☐ Teachers from other districts
- ☐ Master teacher
- ☐ Groups of teachers from the teacher's school
- ☐ State inspector or evaluator
- ☐ School board
- ☐ Students
- ☐ Parents
- ☐ Unspecified
- ☐ Other, please specify _____

6.3 What *expertise and qualifications* are explicitly required of the persons
who evaluate teacher performance?

- ☐ No special qualifications
- ☐ Experience as a teacher
- ☐ Training in administration
- ☐ Experience in administration
- ☐ Training in instructional techniques and methods
- ☐ Training in educational psychology
- ☐ Training in personnel appraisal
- ☐ Knowledge of teaching subject matter
- ☐ Proficiency in particular evaluation methods, please specify _____
- ☐ Knowledge of pedagogy
- ☐ Specialized knowledge of classroom management techniques
- ☐ Specialized knowledge of instructional technique
- ☐ Specialized knowledge of test construction methods
- ☐ Specialized knowledge of classroom grading methods

- ☐ Specialized knowledge of parent involvement techniques
- ☐ Sensitivity to possibilities and risks of linking student learning to teacher performance
- ☐ Knowledge of collegial relationships
- ☐ Sensitivity to and concern for equity
- ☐ Knowledge of the principles and procedures of individual professional development
- ☐ Sensitivity to the influences of the work environment on teaching performance
- ☐ Unspecified
- ☐ Other, please specify _____

7. Evaluation Variables

7.1 What, if any, major categories of entry level *teacher qualifications* are included in the teacher evaluation system?

- ☐ Character traits
- ☐ Morality
- ☐ Attitudes
- ☐ Law abiding
- ☐ General ability
- ☐ Reading skills
- ☐ Writing skills
- ☐ Mathematics skills
- ☐ Speaking skills
- ☐ Listening skills
- ☐ General knowledge
- ☐ Knowledge of field of special competence
- ☐ Knowledge of pervasive curriculum subjects
- ☐ Knowledge of the profession of teaching
- ☐ General pedagogy
- ☐ Designing lessons
- ☐ Subject matter specific pedagogy
- ☐ Ability to generalize and particularize
- ☐ Ability to impart knowledge

- ☐ Involvement in professional association activities
- ☐ Involvement in professional activities
- ☐ Scholarship (knowledge of the professional literature)
- ☐ Caring attitudes toward students
- ☐ Organizational ability (tasking, scheduling, assigning and communicating work plans)
- ☐ Classroom management skills
- ☐ Command of instructional techniques
- ☐ Orientation to service students with special needs
- ☐ Concern for equity
- ☐ Realistic recognition of one's limitations and strengths
- ☐ Commitment to equality of educational opportunity
- ☐ Proficiency in evaluating student performance
- ☐ Proficiency in evaluating classroom activities
- ☐ Physical and emotional stamina to withstand the strains of teaching
- ☐ persistence in sustaining trial and error efforts to solve problems
- ☐ Orientation to serve student needs even if rules need to be bent or broken
- ☐ Awareness and constructive approach to the avoidance of stress and "burn out"
- ☐ Other, please specify _____

7.2 Which of the following *teacher performance criteria* are included in the teacher evaluation system?

- ☐ Ethical conduct
- ☐ Equitable treatment of students and colleagues
- ☐ Professional attitude and performance
- ☐ Knowledge of teaching responsibility
- ☐ Knowledge of school in its context
- ☐ Scholarship (reads the professional literature)
- ☐ Rapport with students
- ☐ Motivation of students
- ☐ Diagnosis of and response to student needs
- ☐ Planning and organization of instruction
- ☐ Supervision of classroom aides

- ☐ Structuring the work of substitute teachers
 - ☐ Involving parents in the education of their children
 - ☐ Classroom management and discipline
 - ☐ Knowledge of field of special competence
 - ☐ Knowledge of pervasive curriculum subjects
 - ☐ Playground management and discipline
 - ☐ Enforcement of school rules
 - ☐ Effectiveness in communicating course content
 - ☐ Command of instructional technology
 - ☐ Demonstrated impact on student achievement
 - ☐ Course development and/or improvement
 - ☐ Course evaluation
 - ☐ Student test scores
 - ☐ Other student performance
 - ☐ Assistance to students with special needs
 - ☐ Individualized assistance to students
 - ☐ Promotion and modeling of equity
 - ☐ Evaluation of student performance
 - ☐ Test construction
 - ☐ Testing
 - ☐ Grading
 - ☐ Reporting student progress
 - ☐ Evaluation and improvement of classroom activities
 - ☐ Personal behavior
 - ☐ Observed strengths
 - ☐ Observed weaknesses
 - ☐ Physical and emotional stamina to withstand the strains of teaching
 - ☐ Compliance with school rules and regulations
 - ☐ Professional development activities
 - ☐ Student judgments of instruction
 - ☐ Cooperation with other school personnel
 - ☐ Global assessment of teaching performance
 - ☐ Other, please specify _____
-

7.3 What, if any, *work environment variables* are assessed and considered in evaluating teacher performance?

- ☐ Availability of appropriate instructional facilities (e.g., photocopy, AV, accessible library)
- ☐ Availability of appropriate instructional materials
- ☐ A safe and drug-free school environment
- ☐ Adequate air conditioning and heating
- ☐ School climate (cooperative atmosphere, orientation to learning, concern for equity)
- ☐ Supportive competent school leadership
- ☐ Adequacy and appropriateness of incentives for excellent teaching
- ☐ Community expectations
- ☐ School's balanced consideration of athletics
- ☐ Family support of student learning
- ☐ School's commitment to academic achievement
- ☐ Students' characteristics, including SES, aptitude, English proficiency, etc.
- ☐ Availability of pedagogical guidance and advice
- ☐ Adequacy and appropriateness of school rules
- ☐ Influence of teacher union or other association
- ☐ Other, please specify _____

8. *Measurement of Performance*

8.1 Which, if any, of the following tools and techniques are used to assess *teacher qualifications*?

- ☐ Basic skills test
- ☐ General knowledge test
- ☐ Knowledge of course content test
- ☐ Pedagogy test
- ☐ Review of credentials
- ☐ Portfolio of teacher's work
- ☐ Videotape of instruction
- ☐ Personality test

- ☐ Job interview
- ☐ Interviews with references
- ☐ Assessment center
- ☐ Simulation exercises
- ☐ Teaching during a trial or probationary period
- ☐ Teaching certificate
- ☐ Continuing Education Units
- ☐ Other, please specify _____

8.2 Which of the following tools and techniques are used to assess *teacher performance*?

- ☐ Principal ratings
- ☐ Student questionnaires
- ☐ Informal observation
- ☐ Videotape of instruction
- ☐ Videotape of student performance
- ☐ Portfolio of teacher performance
- ☐ Portfolio of student performance
- ☐ Classroom observation form
- ☐ Interviewing the teacher
- ☐ Peer observation and coaching
- ☐ Student test scores
- ☐ Parent ratings
- ☐ Other, please specify _____

8.3 Which of the following *rating categories* are used to classify teacher performance (*check all that apply*)?

- ☐ Poor
- ☐ Fair
- ☐ Satisfactory
- ☐ Good
- ☐ Excellent
- ☐ Superior
- ☐ Improvement needed
- ☐ Other, please specify _____

8.4 Which of the following classroom *observation practices* are used in the teacher evaluation system (check all that apply)?

- ☐ Always scheduled in advance
- ☐ Always unannounced
- ☐ Not scheduled in advance
- ☐ Sometimes scheduled in advance
- ☐ No observations conducted

9. Evaluation Reports and Feedback

9.1 Which, if any, of the following *contents* are typically included in the evaluation reports (check all that apply)?

- ☐ List of ratings for various criteria
- ☐ Conference summary
- ☐ Rating of overall effectiveness
- ☐ Narrative assessment of overall effectiveness
- ☐ List of strengths
- ☐ List of weaknesses
- ☐ Recommendations for improvement
- ☐ Timetable for improvement
- ☐ Recommendation on employment status (e.g., continued probation, termination, tenure)
- ☐ Description of data on which the evaluation is based
- ☐ Description of the data collection procedures
- ☐ Other, please specify _____

9.2 Which, if any, of the following steps are included in the evaluation system's *reporting process* (check all that apply)?

- ☐ Evaluatees may review the raw data
- ☐ Evaluator and teacher jointly review the draft report
- ☐ Evaluatee receives final written evaluation report
- ☐ Evaluatee receives a verbal explanation of the written evaluation report
- ☐ Other, please specify _____

9.3 Which, if any, of the following does the evaluation system provide for *attesting the soundness* of evaluation reports?

- ☐ There is an appeal process for evaluations
- ☐ Teacher may signify agreement or disagreement with the report
- ☐ Teacher must signify only to having seen the evaluation report
- ☐ Teacher signs *all* copies of the evaluation report
- ☐ Teacher may attach a written response to the evaluation that becomes a part of the permanent file
- ☐ Other, please specify _____

9.4 Which, if any, of the following apply to the evaluation system's provisions for *distributing evaluation reports* (check all that apply)?

- ☐ A copy of the report is sent to the superintendent's office
- ☐ A copy of the report is provided to the teacher
- ☐ A copy of the report is placed in the school principal's file
- ☐ Filed reports may be accessed by the teacher
- ☐ Filed reports may be accessed by all of the teacher's administrators
- ☐ The teacher sees all copies/versions of the evaluation report
- ☐ Filed reports may be accessed by school board members
- ☐ Other, please specify _____

9.5 Which, if any, of the following are included in the evaluation system's *postobservation review* conferences (check all that apply)?

- ☐ Review satisfactory ratings
- ☐ Review unsatisfactory ratings
- ☐ Give specific suggestions
- ☐ Specify dates for improving deficiencies
- ☐ Schedule a future observation
- ☐ Have teacher acknowledge the conference feedback in writing
- ☐ Provide opportunity for teacher to append a written response
- ☐ Other, please specify _____

10. Use of Evaluation Findings

10.1 How is the evaluation used concerning **individual teachers** (check all that apply)?

- ☐ Teacher is engaged in both a preobservation and postobservation review conference
- ☐ Teacher is engaged only in a postobservation review conference
- ☐ Teacher is engaged only in a preobservation conference
- ☐ School provides guidance for improvements
- ☐ Teacher has the opportunity to design a plan for personal development following evaluation
- ☐ Principal observes/reports implementation of improvements
- ☐ Other, please specify _____

10.2 How are the evaluations used concerning **groups of teachers** (check all that apply)?

- ☐ Not at all
- ☐ Develop district policy
- ☐ Improve supervision
- ☐ Design inservice education
- ☐ Improve selection procedures
- ☐ Change curriculum
- ☐ Change budget allocations
- ☐ Other, please specify _____

10.3 How does the school or school district **remediate/eliminate** deficient performance (check all that apply)?

- ☐ Counseling
- ☐ Professional development activities
- ☐ Specific directives/suggestions
- ☐ Deadlines for improving deficient ratings
- ☐ Extension of the probationary period
- ☐ Termination if remediation efforts fail
- ☐ Unspecified
- ☐ Other, please specify _____

11. Monitoring the Evaluation System—Metaevaluation

11.1 Which, if any, of the following provisions does the district/school employ for *evaluating and improving* the evaluation system?

- ☐ Adherence to the Joint Committee *Personnel Evaluation Standards*
- ☐ Adherence to the APA *Standards for Educational and Psychological Tests*
- ☐ Adherence to the *Equal Employment Opportunity Commission Guidelines*
- ☐ Provision for periodic formal reviews and updating of the evaluation purposes and procedures
- ☐ Annual reviews of the evaluation system
- ☐ Occasional, unscheduled review of the system
- ☐ Reviews if and when the system is challenged
- ☐ External reviews
- ☐ Reliability and validity of the measurement tools have been tested
- ☐ Input from evaluatees is regularly obtained and reviewed
- ☐ System is periodically revised
- ☐ System instruments are periodically reviewed and updated
- ☐ Other, please specify _____

12. Special Provisions

12.1 Which, if any, of the following *groups* in the school or school district are explicitly excluded from the evaluation system reviewed above?

- ☐ Tenured teachers
- ☐ Probationary teachers
- ☐ Art teachers
- ☐ Music teachers
- ☐ Physical education teachers
- ☐ Substitute teachers
- ☐ Special education teachers
- ☐ Classroom aides
- ☐ Unspecified

- ☐ Special support personnel
- ☐ Other, please specify _____

13. Evaluation Models

- 13.1 Which, if any, of the following teacher *evaluation models* or approaches provides the theoretical or logical basis for the teacher evaluation system (*check all that apply*)?

(INSTRUCTIONAL IMPROVEMENT ORIENTED MODELS/
APPROACHES)

- ☐ Madeline Hunter's Instructional Theory Into Practice (ITIP)
- ☐ Richard Manatt's "Clinical Supervision" model
- ☐ Edward Iwanicki's Professional Growth Oriented model
- ☐ Thomas McGreal's Eclectic Professional Development Approach
- ☐ Flanders' Classroom Interaction Model
- ☐ EPIC Classroom Interaction Model (with videotape feedback)
- ☐ Assessment Center approach
- ☐ Micro-teaching
- ☐ Deming—team joint problem-solving approach
- ☐ Other, please specify _____

(PROFESSIONAL ACCOUNTABILITY-DRIVEN MODELS/
APPROACHES)

- ☐ Teacher self-evaluation, a la Tom Good
- ☐ Higher education-type portfolio evaluations
- ☐ Toledo Peer Evaluation Model
- ☐ Peer evaluation (not necessarily patterned after the Toledo model)
- ☐ Resume updates and reviews
- ☐ Professional specialty boards, e.g., National Board for Professional Teaching Standards
- ☐ Other, please specify _____

(ADMINISTRATIVE CONTROL-ORIENTED MODELS/
APPROACHES)

- ☐ Unstructured classroom observation by principal
- ☐ Structured classroom observation by principal
- ☐ Interview/discussion by principal/supervisor or evaluation team
- ☐ Job description-based performance review by principal/supervisor
- ☐ Management by Objectives planning and review by principal and teacher
- ☐ Fitness reports by principal/supervisor, e.g., the military procedure
- ☐ Other, please specify _____

(COLLABORATIVE MODELS/APPROACHES)

- ☐ Anthony Shinkfield's Joint evaluation by principal and peer teachers
- ☐ Other, please specify _____

(RESEARCH-BASED MODELS/APPROACHES)

- ☐ Correlational research-based, structured observation of teacher performance by trained observers
- ☐ Medley, Coker, and Soar—measurement-based teacher evaluation
- ☐ Competency tests
- ☐ Other, please specify _____

(CONSUMER-ORIENTED/COMMUNITY ACCOUNTABILITY
MODELS/APPROACHES)

- ☐ Scriven's Duties-Based Evaluation
- ☐ Parent assessments
- ☐ Student ratings of instruction
- ☐ Student test scores
- ☐ Student test scores corrected for student characteristics
- ☐ Student work products
- ☐ On-site teacher evaluation by governmental department of education inspectors
- ☐ Team visits, managed by state, school district, or other authority

☐ Other, please specify _____

(MERIT PAY MODELS/APPROACHES)

- ☐ Merit increments only, decided by principal/supervisor
- ☐ Merit increments only, decided by peers
- ☐ Merit "bonuses," decided by principal/supervisor
- ☐ Merit "bonuses," decided by peers
- ☐ State-administered Tennessee-type career ladder evaluation approach
- ☐ School/district-administered Tennessee-type career ladder evaluation
- ☐ Merit school approach (no assessment of individual teachers)
- ☐ Other, please specify _____

UNSPECIFIED

- ☐ Not clear that any theoretical approach guides the evaluations

Questions to Be Answered in Addressing the Personnel Evaluation Standards

This appendix is provided for more precise application of the Personnel Evaluation Standards. It poses questions to guide the improvement team to document the degree to which the teacher evaluation system meets individual standards based on the team's response to questions listed under each of the 21 Standard statements. Evidence found in PRINT and PRACTICE should be used to answer these questions.

Standard P-1: Service Orientation

- P-1: Evaluation of educators should promote sound education principles, fulfillment of institutional missions, and effective performance of job responsibilities, so that educational needs of students, community, and society are met.

Questions about your evaluation system relative to the Standard P-1.

	Evidence Found in Print	Evidence Found in Practice
1. Are there provisions for all teachers to be evaluated?	☐ Yes ☐ No	☐ Yes ☐ No
2. Are there provisions for making employment decisions based on evaluation results (e.g., promotion, tenure, remediation, notice to remedy, termination, etc.)?	☐ Yes ☐ No	☐ Yes ☐ No
3. Are there provisions for rewarding outstanding teaching?	☐ Yes ☐ No	☐ Yes ☐ No
4. Are there provisions for evaluating teachers based on differences related to subject, grade level, professional certification, and status in the system, such as probationary, tenure, continuing status?	☐ Yes ☐ No	☐ Yes ☐ No
5. Are there provisions for evaluating how the teacher promotes equitable service to students?	☐ Yes ☐ No	☐ Yes ☐ No
6. Are there provisions for using teacher evaluation results as a basis for designing and implementing specific inservice programs for individual teachers?	☐ Yes ☐ No	☐ Yes ☐ No

- | | | |
|--|--|--|
| 7. Are there provisions for both remediation of deficient performance and step-by-step termination? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 8. Are there provisions for determining whether teachers keep current in their teaching field or other service area? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 9. Do teacher performance criteria include measures of impact on student learning? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 10. Do performance criteria include the overall needs of the students and priorities of the community? | ☐ Yes ☐ No | ☐ Yes ☐ No |

Standard P-2: Formal Evaluation Guidelines

- P-2: Guidelines for personnel evaluations should be reported in statements of policy, negotiated agreements, and/or personnel evaluation manuals, so that evaluations are consistent, equitable, and in accordance with pertinent laws and ethical codes.

Questions about your evaluation system relative to the Standard P-2.

- | | Evidence
Found in
Print | Evidence
Found in
Practice |
|--|--|--|
| 1. Are there guidelines for implementing the evaluation procedures contained in policies, negotiated agreements, and/or personnel evaluation manuals? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 2. Are the evaluation criteria limited to important job-related issues? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 3. Are both guidelines for implementation of evaluation policy and evaluation criteria clear, specific, and understandable? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 4. Are there provisions in policies, negotiated agreements, and/or evaluation manuals for appropriate emphasis (weights) to be assigned each evaluation criterion before it is applied? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 5. Are there provisions to assure that local, state, and federal requirements—such as state tenure laws, teacher certification laws, equity laws, and other guidelines—are adhered to in employment decisions? | ☐ Yes ☐ No | ☐ Yes ☐ No |

- | | | |
|--|--|--|
| 6. Are there provisions for explaining the evaluation system and its application to all evaluatees annually and at times in between when changes occur? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 7. Are there provisions for implementing remediation plans in progressive stages? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 8. Are there clear and precise statements that define types of evaluation findings likely to lead to termination? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 9. Are there provisions for changing formal evaluation guidelines when evaluation practices are changed, when guidelines are in conflict with laws, or when role definitions change? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 10. Are there guidelines governing both the frequency of evaluations and a time line for implementing evaluation stages? | ☐ Yes ☐ No | ☐ Yes ☐ No |

Standard P-3: Conflict of Interest

- P-3:** Conflicts of interest should be identified and dealt with openly and honestly, so that they do not compromise the evaluation process and results.

Questions about your evaluation system relative to the Standard P-3.

- | | Evidence
Found in
Print | Evidence
Found in
Practice |
|--|--|--|
| 1. Are there provisions for cooperation among the district governing board, administrators, teachers, and other stakeholder groups in designing the evaluation system? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 2. Are there provisions for identifying and documenting common sources of conflicts of interest in the evaluation system and its application? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 3. Are there provisions for controlling conflicts of interest as part of the selection of personnel who will conduct evaluations? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 4. Are there provisions for use of clear criteria and objective evidence where indicated as a basis for evaluation? | ☐ Yes ☐ No | ☐ Yes ☐ No |

- | | | |
|--|--|--|
| 5. Are there provisions for involvement of the evaluatee in the review of the process and resulting evidence before finalizing the evaluation report? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 6. Are there provisions that clearly designate which evaluation findings may be used in the event of appeal? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 7. Does the evaluation system provide for the use of multiple sources of information, such as self-evaluation, evaluation by students, evaluation by peers, observation, portfolios, etc.? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 8. Are there provisions for designating an alternate evaluator or evaluators if an unresolvable conflict exists? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 9. Are there provisions for reaching agreement between the evaluator and the evaluatee on the criteria to be used in assessing performance and the conditions under which the evaluation is to take place? | ☐ Yes ☐ No | ☐ Yes ☐ No |

Standard P-4: Access to Personnel Evaluation Reports

- P-4: Access to reports of personnel evaluation should be limited to individuals with a legitimate need to review and use the reports, so that appropriate use of the information is assured.

Questions about your evaluation system relative to the Standard P-4.

- | | Evidence
Found in
Print | Evidence
Found in
Practice |
|--|--|--|
| 1. Are there provisions for secure storage of evaluation information collected prior to final reports? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 2. Are there provisions for identifying who shall have access to evaluation reports and when and why they shall have access? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 3. Are there provisions for the basis and procedures for removing evaluation information from the school or central files? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 4. Are there provisions for deleting and adding to personnel evaluation reports? | ☐ Yes ☐ No | ☐ Yes ☐ No |

- | | | | | |
|---|------------------------------|-----------------------------|------------------------------|-----------------------------|
| 5. Are there provisions for secure storage of both manual and electronic evaluation reports and other related records? | ☐ Yes | ☐ No | ☐ Yes | ☐ No |
| 6. Are there provisions specifying who will receive copies of the report? | ☐ Yes | ☐ No | ☐ Yes | ☐ No |
| 7. Are there provisions for the evaluatee to receive a signed copy of the final evaluation report, including any appendices? | ☐ Yes | ☐ No | ☐ Yes | ☐ No |
| 8. Are there provisions for discussing all information with the evaluatee before it is placed in the official personnel file? | ☐ Yes | ☐ No | ☐ Yes | ☐ No |
| 9. Are there provisions for limiting access to reports to those who must make or defend decisions based on them and to those designated in writing by the employee? | ☐ Yes | ☐ No | ☐ Yes | ☐ No |
| 10. Is training in release and retrieval of evaluation information provided for those who have access to and use records in personnel files? | ☐ Yes | ☐ No | ☐ Yes | ☐ No |

Standard P-5: Interaction with Evaluatees

- P-5: The evaluation should address evaluatees in a professional, considerate, and courteous manner, so that their self-esteem, motivation, professional reputations, performance, and attitude toward personnel evaluation are enhanced or, at least, not needlessly damaged.

Questions about your evaluation system relative to the Standard P-5.

- | | Evidence
Found in
Print | Evidence
Found in
Practice |
|---|--|--|
| 1. Are there timetables that guide evaluation stages? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 2. Are there provisions for setting specific evaluation timetable dates in cooperation with evaluatees? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 3. Are there provisions for setting and conforming to stated performance goals and objectives that are mutually agreed on by the evaluator and the evaluatee? | ☐ Yes ☐ No | ☐ Yes ☐ No |

- | | | |
|---|--|--|
| 4. Are there provisions for immediate assistance or intervention when performance deficiencies require such response? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 5. Are there provisions for encouraging and assisting professional growth? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 6. Are there provisions for providing review and feedback on strengths and weaknesses of performance in private uninterrupted sessions? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 7. Are there provisions for an appeal process for evaluations? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 8. Are there provisions for evaluatees to signify agreement or disagreement with the evaluation report and append written response? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 9. Are there provisions for evaluatees to receive a copy of the final evaluation report? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 10. Are there provisions for requiring evaluators to receive training in human interaction? | ☐ Yes ☐ No | ☐ Yes ☐ No |

Standard U-1: Constructive Orientation

- U-1: Evaluations should be constructive, so that they help institutions to develop human resources and encourage and assist those evaluated to provide excellent service.

Questions about your evaluation system relative to the Standard U-1.

- | | Evidence
Found in
Print | Evidence
Found in
Practice |
|--|--|--|
| 1. Are there provisions for the district governing board to formally adopt the teacher evaluation system? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 2. Are there provisions for representation of all stakeholders in defining performance standards? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 3. Are there provisions for representation of all stakeholders in defining respective roles in evaluating teachers, e.g., principals, peers, students, evaluatees, others? | ☐ Yes ☐ No | ☐ Yes ☐ No |

- | | | |
|--|--|--|
| 4. Are there provisions for communicating to all stakeholders the importance of teacher evaluation for professional development and the achievement of organizational goals? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 5. Are there provisions for beginning evaluation conferences with positive communication, e.g., performance strengths? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 6. Are there provisions for emphasizing support for the teacher as a professional (e.g., funds for additional training and additional coursework, released time for collaboration with colleagues or consultants)? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 7. Are there provisions for identifying performance areas that require reinforcement and/or improvement? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 8. Are there provisions for specific written directives and recommendations for remediation of deficient performance? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 9. Are there provisions for providing resources for improving performance (e.g., assistance from master teachers, instructional leaders, and/or funds for materials)? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 10. Are there provisions for encouraging and assisting teachers in assessing and improving their own performance? | ☐ Yes ☐ No | ☐ Yes ☐ No |

Standard U-2: Defined Uses

- U-2: The users and the intended uses of a personnel evaluation should be identified, so that the evaluation can address appropriate questions.

Questions about your evaluation system relative to the Standard U-2.

- | | Evidence
Found in
Print | Evidence
Found in
Practice |
|--|--|--|
| 1. Are there provisions for identifying and informing all potential audiences of the content and availability of evaluation reports? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 2. Are there provisions for evaluatees to learn of the intended audiences of evaluation reports and results.? | ☐ Yes ☐ No | ☐ Yes ☐ No |

3. Are there provisions for constructing evaluation inquiries that are relevant to information needs and proposed uses? ☐ Yes ☐ No ☐ Yes ☐ No
4. Are there provisions for limiting audiences to, and uses for, evaluation reports to those mutually agreed on prior to the evaluation cycle? ☐ Yes ☐ No ☐ Yes ☐ No

Standard U-3: Evaluator Credibility

U-3: The evaluation system should be managed and executed by persons with the necessary qualification, skills, and authority. And evaluators should conduct themselves professionally, so that evaluation reports are respected and used.

Questions about your evaluation system relative to the Standard U-3.

- | | Evidence
Found in
Print | Evidence
Found in
Practice |
|--|--|--|
| 1. Are there provisions for requiring evaluators to be knowledgeable about each of the following: a variety of sound teaching techniques, the principles of learning psychology, and the implications of human growth and development for effective teaching? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 2. Are there provisions for training district governing board members, administrators, faculty, and evaluation specialists for maximum effectiveness in their evaluation roles? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 3. Are there provisions requiring those who serve as evaluators to become knowledgeable in principles of sound personnel evaluation, performance appraisal techniques, methods of motivating faculties, conflict management, and the law as it applies to evaluation of educational personnel? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 4. Are there provisions for establishing the authority and responsibilities of evaluators? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 5. Are there provisions for more than one evaluator to be involved in gathering information about an individual teacher? | ☐ Yes ☐ No | ☐ Yes ☐ No |

- | | | | | |
|--|------------------------------|-----------------------------|------------------------------|-----------------------------|
| 6. Are there provisions for adding resources to assist in information collection and analysis when the tasks exceed the professional competence of evaluators? | ☐ Yes | ☐ No | ☐ Yes | ☐ No |
| 7. Are there provisions for maintaining the same evaluator(s) throughout any single evaluation? | ☐ Yes | ☐ No | ☐ Yes | ☐ No |
| 8. Are there provisions for the preparation and use of a relevant agenda (shared in advance with the evaluatee) during feedback sessions? | ☐ Yes | ☐ No | ☐ Yes | ☐ No |

Standard U-4: Functional Reporting

- U-4: Reports should be clear timely, accurate, and germane, so that they are of practical value to the evaluatee and other appropriate audiences.

Questions about your evaluation system relative to the Standard U-4.

- | | Evidence
Found in
Print | Evidence
Found in
Practice |
|---|--|--|
| 1. Are there provisions requiring that multiple criteria be used in evaluating teaching performance? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 2. Are there provisions for requiring a rating of overall effectiveness of teaching performance? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 3. Are there provisions for a timetable for professional growth? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 4. Are there provisions for including evaluation information in recommendations determining employment status (i.e., continued probation, termination, tenure, or continued service)? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 5. Are there provisions for initiating evaluations early enough in the school year to allow time for interim reporting? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 6. Are there provisions for addressing <i>only</i> identified and agreed-on professional responsibilities in the evaluation report? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 7. Are there provisions for prompt written reports to be given to the evaluatee by evaluators following formal observation of an evaluatee? | ☐ Yes ☐ No | ☐ Yes ☐ No |

Standard U-5: Follow-Up and Impact

- U-5: Evaluations should be followed up, so that users and evaluatees are aided to understand the results and appropriate actions.

Questions about your evaluation system relative to the Standard U-5.

	Evidence Found in Print	Evidence Found in Practice
1. Are there provisions for reviewing performance strengths and weaknesses with the evaluatee and soliciting suggestions for improvement?	☐ Yes ☐ No	☐ Yes ☐ No
2. Are there provisions for assisting in improving identified performance weaknesses and establishing a plan for improvement?	☐ Yes ☐ No	☐ Yes ☐ No
3. Are there provisions for holding follow-up conferences between the evaluatee and appropriate resource personnel when such conferences are necessary?	☐ Yes ☐ No	☐ Yes ☐ No
4. Are there provisions for flexibility in planning, with evaluatee input, for professional growth to reinforce strengths and overcome identified weaknesses?	☐ Yes ☐ No	☐ Yes ☐ No
5. Are there provisions to assist the evaluatee with resources, released time, and/or other action to assure that the professional growth plan will succeed?	☐ Yes ☐ No	☐ Yes ☐ No
6. Are there provisions for non-reemployment notices to be given by a specified appropriate date?	☐ Yes ☐ No	☐ Yes ☐ No
7. Are there provisions for scheduling the next evaluation or evaluation stage during the follow-up conference?	☐ Yes ☐ No	☐ Yes ☐ No
8. Are there provisions for making and keeping written records of follow-up conferences, progress toward agreed-on goals and objectives, and results?	☐ Yes ☐ No	☐ Yes ☐ No
9. Are there provisions to ensure realistic implementation of both remediation and professional growth plans?	☐ Yes ☐ No	☐ Yes ☐ No

10. Are there provisions for follow-up conferences to be held with the evaluatee within a reasonable time following each observation? ☐ Yes ☐ No ☐ Yes ☐ No
11. Are there provisions for the evaluatee to acknowledge or respond in writing to conference feedback? ☐ Yes ☐ No ☐ Yes ☐ No
12. Are there provisions for using evaluation results as an information source in planning curriculum change, designing inservice education, allocating budget funds, developing district policy, and improving supervision? ☐ Yes ☐ No ☐ Yes ☐ No

Standard F-1: Practical Procedures

- F-1: Personnel evaluation procedures should be planned and conducted so that they produce needed information while minimizing disruption and cost.

Questions about your evaluation system relative to the Standard F-1.

- | | Evidence
Found in
Print | Evidence
Found in
Practice |
|---|--|--|
| 1. Are there provisions that information collection will be determined, modified, and applied with minimum disruption? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 2. Are there provisions for identifying needs, available resources, and policy requirements in designing, selecting, and improving information collection procedures? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 3. Are there provisions for avoiding or eliminating the duplication of evaluation information that already exists? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 4. Are there provisions for periodic orientation sessions to help educators understand the purposes and processes of the evaluation system? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 5. Are there provisions for encouraging teachers and other stakeholders to suggest ways by which evaluation procedures can be made more useful? | ☐ Yes ☐ No | ☐ Yes ☐ No |

6. Are there provisions for limiting the collection of evaluation information to that which is relevant to the position and the purposes of the evaluation?

☐ Yes ☐ No ☐ Yes ☐ No

Standard F-2: Political Viability

- F-2: The personnel evaluation system should be developed and monitored collaboratively, so that all concerned parties are constructively involved in making the system work.

Questions about your evaluation system relative to the Standard F-2.

1. Are there provisions requiring that policies established by the district governing board become final authority in determining evaluation matters?
2. Are there provisions for a continuing and representative improvement team to periodically develop, revise, and propose evaluation policy?
3. Are there provisions for promptly and effectively addressing problems in the personnel evaluation system?
4. Are there provisions for informing teachers and other stakeholders of the evaluators' responsibilities?
5. Are there provisions for arriving at mutual agreement between the policy board and school staff on evaluation policy and procedures?
6. Are there provisions for informing stakeholders of agreed-on evaluation policy and procedures (e.g., through newsletters, open meetings, board minutes, etc.)?

Evidence Found in Print

Evidence Found in Practice

☐ Yes ☐ No ☐ Yes ☐ No

☐ Yes ☐ No ☐ Yes ☐ No

☐ Yes ☐ No ☐ Yes ☐ No

☐ Yes ☐ No ☐ Yes ☐ No

☐ Yes ☐ No ☐ Yes ☐ No

☐ Yes ☐ No ☐ Yes ☐ No

Standard F-3: Fiscal Viability

F-3: Fiscal Viability: Adequate time and resources should be provided for personnel activities, so that evaluation plans can be effectively and efficiently implemented.

*Questions about your evaluation system
relative to the Standard F-3.*

	Evidence Found in Print	Evidence Found in Practice
1. Are there provisions for sufficient allocations of resources to meet the defined purposes, procedures, and uses of results?	☐ Yes ☐ No	☐ Yes ☐ No
2. Are there provisions for a minimum of procedures and time to be expended in obtaining the needed information?	☐ Yes ☐ No	☐ Yes ☐ No
3. Are there provisions for allocation of staff time and frequency of evaluations based on reasonable estimates of the time required to conduct each type of evaluation?	☐ Yes ☐ No	☐ Yes ☐ No
4. Are there provisions for funds to carry out the procedures mandated?	☐ Yes ☐ No	☐ Yes ☐ No
5. Are there provisions for monitoring the efficiency and effectiveness of the system (evaluation of the evaluation)?	☐ Yes ☐ No	☐ Yes ☐ No
6. Are there provisions for a continuous search for new ideas that will result in achieving and maintaining the highest possible cost effectiveness of the evaluation system?	☐ Yes ☐ No	☐ Yes ☐ No

Standard A-1: Defined Role

- A-1: The role, responsibilities, performance objectives, and needed qualifications of the evaluatee should be clearly defined, so that the evaluator can determine valid assessment data.

*Questions about your evaluation system
relative to the Standard A-1.*

	Evidence Found in Print	Evidence Found in Practice
1. Are there provisions for position descriptions that clearly delineate educational assignment (e.g., grade level, subject area, special program areas, etc.)?	☐ Yes ☐ No	☐ Yes ☐ No
2. Are there provisions for evaluating important responsibilities that are other than instructional (i.e., work habits, cooperation with colleagues, and so forth)?	☐ Yes ☐ No	☐ Yes ☐ No
3. Are there provisions for evaluating entrance qualifications for special fields of expertise or teaching areas when the teaching area is changed?	☐ Yes ☐ No	☐ Yes ☐ No
4. Are there provisions for internal notification (within the school) and external communication (within the district) of both performance criteria and the level of performance acceptable in the school district?	☐ Yes ☐ No	☐ Yes ☐ No
5. Are there provisions for periodic reviewing and updating of performance criteria and job descriptions?	☐ Yes ☐ No	☐ Yes ☐ No
6. Are there provisions that require proficiency of evaluatees in assessing, recording, and reporting student performance?	☐ Yes ☐ No	☐ Yes ☐ No
7. Are there provisions for determining the level of evaluatees' involvement in professional association activities?	☐ Yes ☐ No	☐ Yes ☐ No
8. Are there provisions for assessing teachers' knowledge of other curriculum areas that are relevant to their teaching assignment?	☐ Yes ☐ No	☐ Yes ☐ No

- | | | | | |
|---|------------------------------|-----------------------------|------------------------------|-----------------------------|
| 9. Are there provisions for assessing teachers' understanding of the specific contribution to be made to the overall curriculum by their particular assigned teaching position? | ☐ Yes | ☐ No | ☐ Yes | ☐ No |
| 10. Are there provisions for assessing whether or not students receive fair treatment by teachers? | ☐ Yes | ☐ No | ☐ Yes | ☐ No |
| 11. Are there provisions for investigating and resolving conflicting or inaccurate provisions within position descriptions? | ☐ Yes | ☐ No | ☐ Yes | ☐ No |

Standard A-2: Work Environment

A-2: The context in which the evaluatee works should be identified, described, and recorded, so that environmental influences and constraints on performance can be considered in the evaluation.

Questions about your evaluation system relative to the Standard A-2.

- | | Evidence
Found in
Print | Evidence
Found in
Practice |
|--|--|--|
| 1. Are there provisions for considering and recording the availability and appropriateness of instructional facilities and materials (e.g., photocopiers, AV equipment, accessible library, texts, and other instructional media and materials)? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 2. Are there provisions for considering and recording the condition of the building, room, or other facility in which the performance is being assessed? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 3. Are there provisions for considering and recording availability of professional, paraprofessional, and secretarial support services to the teacher? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 4. Are there provisions for considering and recording student characteristics as they affect teacher performance? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 5. Are there provisions for considering the adequacy and appropriateness of school rules and regulations as they affect teacher performance? | ☐ Yes ☐ No | ☐ Yes ☐ No |

6. Are there provisions for considering in the evaluation the number of students the teacher must work with during the day?

☐ Yes ☐ No ☐ Yes ☐ No

Standard A-3: Documentation of Procedures

- A-3: The evaluation procedures actually followed should be documented, so that the evaluatee and other users can assess the actual, in relation to intended, procedures.

Questions about your evaluation system relative to the Standard A-3.

Evidence Found in Print

Evidence Found in Practice

1. Are there provisions for the use of a district-governing-board-approved evaluation procedure?
2. Are there provisions for the use of district-governing-board-approved evaluation forms?
3. Are there provisions for recording performance ratings based on established criteria?
4. Are there provisions for keeping written records of conferences with individual evaluatees associated with performance evaluation?
5. Are there provisions for including all sources of evaluation data in evaluation reports?
6. Are there provisions for informing evaluatees in writing of the established procedures?

☐ Yes ☐ No ☐ Yes ☐ No

☐ Yes ☐ No ☐ Yes ☐ No

☐ Yes ☐ No ☐ Yes ☐ No

☐ Yes ☐ No ☐ Yes ☐ No

☐ Yes ☐ No ☐ Yes ☐ No

☐ Yes ☐ No ☐ Yes ☐ No

Standard A-4: Valid Measurement

- A-4: The measurement procedures should be chosen or developed and implemented on the basis of the described role and the intended use, so that the inferences concerning the evaluatee are valid and accurate.

*Questions about your evaluation system
relative to the Standard A-4.*

	Evidence Found in Print	Evidence Found in Practice
1. Are there provisions for collecting evaluation information from a variety of sources?	☐ Yes ☐ No	☐ Yes ☐ No
2. Are there provisions for ensuring that sources of evaluation information used conform with evaluation system guidelines?	☐ Yes ☐ No	☐ Yes ☐ No
3. Are there provisions for evaluating performance against clear descriptions of performance criteria?	☐ Yes ☐ No	☐ Yes ☐ No
4. Are there provisions for involving stakeholders in determining the appropriateness of purposes, criteria, processes, and instruments used in evaluation?	☐ Yes ☐ No	☐ Yes ☐ No
5. Are there provisions assuring that agreed-on sequences will be carried out in the evaluation process?	☐ Yes ☐ No	☐ Yes ☐ No
6. Are there provisions for limiting evaluation to assessing agreed-upon performance criteria?	☐ Yes ☐ No	☐ Yes ☐ No
7. Are there provisions for clearly and precisely describing data on which evaluation is based?	☐ Yes ☐ No	☐ Yes ☐ No
8. Are there provisions for assuring that the instruments and processes accurately evaluate the intended system purposes and criteria?	☐ Yes ☐ No	☐ Yes ☐ No

Standard A-5: Reliable Measurement

- A-5: Measurement procedures should be chosen or developed to assure reliability, so that the information obtained will provide consistent indications of the performance of the evaluatee.

Questions about your evaluation system relative to the Standard A-5.

	Evidence Found in Print	Evidence Found in Practice
1. Are there provisions for training observers to apply evaluation criteria consistently and objectively?	☐ Yes ☐ No	☐ Yes ☐ No
2. Are there provisions for training of evaluators in the intended use of procedures and instruments?	☐ Yes ☐ No	☐ Yes ☐ No
3. Are there provisions for testing the consistency of procedures across evaluators and making changes indicated by the findings?	☐ Yes ☐ No	☐ Yes ☐ No
4. Are there provisions for ensuring consistency of instruments throughout the district?	☐ Yes ☐ No	☐ Yes ☐ No
5. Are there provisions for pilot testing changes in procedures and instruments before full implementation to assure their consistency?	☐ Yes ☐ No	☐ Yes ☐ No

Standard A-6: Systematic Data Control

- A-6: The information used in the evaluation should be kept secure, and should be carefully processed and maintained, so as to ensure that the data maintained and analyzed are the same as the data collected.

Questions about your evaluation system relative to the Standard A-6.

	Evidence Found in Print	Evidence Found in Practice
1. Are there provisions for training those who handle and process evaluation information to perform their tasks with appropriate care and discretion?	☐ Yes ☐ No	☐ Yes ☐ No
2. Are there provisions requiring that a sign-out procedure be followed when removing files from storage?	☐ Yes ☐ No	☐ Yes ☐ No

- | | | |
|--|--|--|
| 3. Are there provisions for identifying person/position and reason for addition to or removal of materials from personnel evaluation files? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 4. Are there provisions for maintaining backup files in a secure location? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 5. Are there provisions for requiring evaluation documents to be labeled ORIGINAL or COPY? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 6. Are there provisions for developing and maintaining an appropriate filing system, so that information can be easily and accurately retrieved when needed? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 7. Are there provisions to ensure that files removed from storage locations will be returned in their original form? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 8. Are there provisions for informing evaluatees of the distribution (to whom, when, and why) of evaluation reports? | ☐ Yes ☐ No | ☐ Yes ☐ No |

Standard A-7: Bias Control

- A-7: The evaluation process should provide safeguards against bias, so that the evaluatee's qualifications or performance are assessed fairly.

Questions about your evaluation system relative to the Standard A-7.

- | | Evidence
Found in
Print | Evidence
Found in
Practice |
|---|--|--|
| 1. Are there provisions for prompt third party reviews of appeals? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 2. Are there provisions for monitoring the evaluation process so it will not focus on aspects of performance or personal activities irrelevant to identified roles? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 3. Are there provisions for reporting relevant information even if it conflicts with the general conclusions or recommendations? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 4. Are there provisions for the evaluator and teacher to jointly review the draft evaluation report? | ☐ Yes ☐ No | ☐ Yes ☐ No |

5. Are there provisions for having written feedback from the teacher regarding the teacher/evaluator conference? ☐ Yes ☐ No ☐ Yes ☐ No

Standard A-8: Monitoring Evaluation Systems

- A-8: The personnel evaluation system should be reviewed periodically and systematically, so that appropriate revisions can be made.

Questions about your evaluation system relative to the Standard A-8.

- | | Evidence
Found in
Print | Evidence
Found in
Practice |
|---|--|--|
| 1. Are there provisions for determining the positive effects of teacher evaluation on the results of schooling? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 2. Are there provisions for budgeting sufficient resources and personnel for periodic review of the evaluation system? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 3. Are there provisions for reviewing policies and procedures of evaluation to determine if they are still appropriate and effective? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 4. Are there provisions for comparing evaluation plans to actual practice? | ☐ Yes ☐ No | ☐ Yes ☐ No |
| 5. Are there provisions for periodically surveying staff to obtain critiques and recommendations related to evaluation policies and procedures? | ☐ Yes ☐ No | ☐ Yes ☐ No |

4

MODELS FOR TEACHER EVALUATION

Preamble: Overview of Alternative Models

Introduction

The purpose of this Preamble, and of Part 4, is to give an overview of teacher evaluation models and to present a range of models that either are widely used or are likely to be increasingly influential because of their practical and unique characteristics.

It will be seen that a preponderance of models selected in this chapter emphasize improvements in classroom teaching and, either stated or implied, strengthened student learning. The reason for this is that almost all teacher evaluation systems adopted (or adapted) by school districts have a strong component of teacher professional development. This has already been stated in Part 1 (Historical Perspectives). Some of these kinds of models also include teacher accountability as an aspect of teacher development. Other models presented give more emphasis to administrative control or to administrative responsiveness to community concerns, particularly those associated with demonstrable student learning.

What is Meant by a Model? The introduction chapter has stated our intention to plan this book around four dominant, interrelated cores—professional standards for developing and evaluating evaluation systems, a **GUIDE** for applying the Joint Committee's *Standards*, ten models for evaluating teacher performance, and an

analysis of these selected models. As the term "model" is centrally featured, it requires some definition.

Unlike mathematical models used to test theory, each model presented in Part 4 characterizes the author's view of the main concepts involved in approaching the tasks of teacher evaluation. The models operationalize these concepts by providing guidelines for developing themes and activities to a stage where justifiable conclusions have been underpinned by credible description, advice, and judgments.

It has been contended that the word "model" should be related only to a series of directions leading to designed conclusions and that alternative perspectives should not be accorded the status of model. Like earlier writers on this topic (Madaus et al., 1983), we are satisfied to redirect emphasis away from the characterization of the various conceptualizations of teacher evaluation as models of evaluation carried out in an ordered iteration, to their characterization as models for conducting studies according to the beliefs of the various authors whose work is represented in this book. In this latter sense, there may be offered a conceptual (and sometimes idealized) view of what teacher evaluation should be. A good example of this is Graeme Withers' article in Chapter 4.4 on teacher self-evaluation where theory evolves into concepts, which in turn lead into proposed activities by participants and procedures for monitoring and judging the worth of these activities. Most of the models, however, are directive in purpose and procedure.

Whether the models given in Part 4 are more directive or less directive for the user, they have all been based on a similar intent—to evaluate teachers so well that there are clear benefits for schools and school districts as an outcome of teachers being increasingly aware of their professional responsibilities.

If space had allowed, it would have been possible to refer to other writings to show further various authors' beliefs about evaluation and its potential uses. For instance, Hans Andrews' book, *Evaluation for Excellence* (1985), gives a series of sharp snapshots directed at salient features of summative teacher evaluation. Andrews writes that an evaluation system "must assist faculty members to improve for retention or promotional purposes and also must provide assurances that incompetent faculty can be removed for the best interest of students and public policy" (pp. 19-20). A version of this book has not been included here, as it does not offer a "modeling" approach, but rather a hard-hitting series of advisory statements to guide educational leaders and school board members. By comparison, a depiction of the Toledo School District's Intern and Intervention Programs, which includes the summative role of evaluation, is provided in Chapter 4.6 because it conforms to the way in which a model has been defined earlier in this section.

A further most valuable contribution to teacher evaluation practice is Michael Scriven's duties-based approach (1993). This is founded on criteria derived from a normative study of what teachers legally can be expected to do. The Scriven duties list is somewhat similar to lists developed by Shinkfield in 1982 (see Chapter 4.7).

Iwanicki in 1983 (see Chapter 4.3), and McGreal in 1985 (see Chapter 4.2), but is considerably more developed and sophisticated. The Scriven duties-based approach could be incorporated into existing or new formative or summative models, adding considerable strength. Thus, while it cannot be considered to be a model, it can be considered as a valuable adjunct to the process of teacher evaluation.

An Overview of Teacher Evaluation Models TEMP Memo 2 (September 1991) offers a succinct overview of 15 models – “ways to evaluate teachers that implicitly define good teaching” (p. 6) – which are labeled according to their most distinctive feature. Most of these are outlined below, and one omission from the TEMP Memo, self-evaluation, is added.

The first four models cover classroom observation, which is regularly used for inservice evaluation and less frequently for promotion or merit award reasons. These models may be implemented by the principal, other administrative or educational leaders, or trained teams. It will be seen that classroom observation has remained the dominant element in teacher evaluation.

1. *Traditional Impressionistic*

Judgments are made based on the observers' (usually principals') experience and educational views.

2. *Clinical Supervision*

Madeline Hunter is the chief proponent of this approach, but somewhat similar versions abound.

3. *Research-Based Checklist*

This is the most prevalent of recent approaches, forming the basis of most state-mandated evaluation instruments.

4. *High Inference Judgments*

Evaluators undergo specialized training to help ensure that skilled, reliable judgments occur.

5. *Interviewing*

This probably will be an ongoing process for teacher professional development or for decision making about a teacher's status within a school: it has many potential uses besides selection, including promotion, remediation, reassignment, and potential dismissal.

6. *Paper and Pencil Tests*

These may be used for national teaching examinations, (e.g., the National Board for Professional Teaching Standards); they are far more likely to be used for entry and reclassification purposes than for inservice appraisal.

7. *Management By Objectives*
This focuses on mutually agreed goals and usually has a designated iteration and agreed measures indicating success in meeting objectives. Use is made of portfolios of artifacts as well as classroom observation.
8. *Job Analysis*
This is based on criteria arising from a descriptive examination of what teachers actually do. Observations and other data are basic sources of information. This approach is sometimes allied to *competency-based* teacher evaluation.
9. *Duties-Based Approach*
This has been explained and referred to again in the final paragraph of the previous section of this preamble.
10. *Theory-Based Approach*
This may derive evaluative criteria from a pertinent theory, e.g., a theory that links student achievement to certain teaching practices.
11. *Student Learning (Improvement) Outcomes*
The key variable is the measurement of student learning improvement at designated times during a school year or over a period of years. This approach has gained considerable momentum in recent years (as Part 4 will show).
12. *Consumer Ratings*
Student ratings are common at college level and rare at school level. Parents seldom are requested to rate teachers (at least formally).
13. *Peer Ratings*
This has increased in popularity in recent years. (It forms an integral part of three of the models outlined in the next section.)
14. *Self-Evaluation*
This model is open to criticism (usually based on weak validity and suspect reliability), but formal attempts to strengthen the approach are being made. Chapter 4.4 is one such example.
15. *Metaevaluation of Existing Models*
The main example of stipulated general criteria to judge the work of models is the Joint Committee's *Standards* (1988).

The preamble continues with an overview of ten models. These have been selected because they cover many of the TEMP Memo approaches outlined above. Moreover, they all are either widely used or are influential in that they are recognized for their contribution, often unique, to the advancement of teacher evaluation. If their presentation achieves no other purpose, we believe that they indicate the breadth of the range of models available for readers to consider. In Part 5 (An Analysis of Alternative Models) the models presented in Part 4 are summa-

rized and then contrasted based on three different, but related, ways of viewing them. This should also prove useful for decision-making purposes.

An Overview of Ten Selected Models

This section presents a brief summary of the ten models that comprise Part 4. No attempt will be given here to offer value judgments about each model. Such judgments will be given in Part 5, where the models' main strengths and weaknesses identified by comparison with the Joint Committee's *Standards* will be offered. Although discussed at some length in Part 5, it should be noted here that the first four models mainly comply with the formative role of evaluation, the next three with both formative and summative roles, and the final three with the summative role.

Chapter 4.1 Madeline Hunter: Instructional Effectiveness Through Clinical Supervision

Dr. Madeline Hunter gained an international reputation for promoting studies in the cause-effect relationships between teaching and learning. Her work and writings, and leadership in extensive inservice training and workshop sessions with educators, have emphasized the importance of the teacher as an instructional decision maker and have helped to clarify the artistry of teaching.

The "Hunter Models" had their origins during the early 1970s in the Teacher Appraisal Instrument (TAI), which developed into the Teacher Appraisal Instructional Improvement Instrument (TA Triple I). Through formative evaluation techniques, both instruments focus on observing teachers giving instruction to particular students in particular situations. Concern lies with what a teacher *does*, and not what a teacher *is*. The model has not been created for evaluation purposes, but for increasing teacher excellence. The observer makes decisions about growth-evoking feedback and pinpoints effective decisions, reinforces them and states the principles undergirding them, as well as inappropriate teaching decisions, and offers productive alternatives. Moreover, the observer discusses and advises, not admonishes. To this extent, at least, the model has strong elements of formative evaluation.

In summary, the model

- allows teachers to identify professional decisions they must make
- offers causal relationships (often research-based) to support these decisions
- encourages teachers to use analyzed instructional information to strengthen present practices that are successful, to develop additional productive alternatives, or to correct their decisions so that the probability of learning is increased

- involves district educators with a potential for leadership and intensive planning for successful inservice by both leaders and teachers, which is the forerunner of the implementation of the **Lesson Design**. This consists of seven elements for the planning of effective instruction, focusing on principles of learning so that student learning may be accelerated.

Chapter 4.1 discusses the various components of the Hunter Model. It should be observed, however, that once leaders are trained and the model is about to be implemented, the specific needs of teachers and schools may dictate where time and effort are to be spent to make the approach work successfully. It should also be noted that whereas supervisory conferencing skills are contained in a separate section of Chapter 4.1, the various types of conferences that take place between the teacher and the administrator will be determined by the teacher's sophistication in implementing cause-effect relationships in students' learning.

Unlike some other teacher evaluation models or teacher improvement approaches that have evaluative components, the Hunter program is designed to make all outcomes as productive as possible. In fact, one of the most productive outcomes is the focusing of decision making on the teacher, who is encouraged and even compelled to analyze teaching situations for both learner and teacher enhancement.

Chapter 4.2: Thomas McGreal: Characteristics of Successful Teacher Evaluation

Thomas McGreal has worked with hundreds of school districts over the years to encourage the design and development of realistic and effective systems of teacher evaluation. His main intention is not to advocate one particular approach to evaluation, but to emphasize certain concepts, or "commonalities" as he terms them, that may become the basis for decisions. The fact that many school districts have followed his advice, and continue to do so, indicates his importance in the growth of the teacher evaluation movement.

McGreal states that there are two issues that a school district must address if its present teacher evaluation approach is to improve or if a new one is to be effective. First, congruency must exist between what the school district wants the evaluation system to do and to be and those things that the evaluation approach requires of the personnel involved. Second, because evaluations necessarily lead to decisions, McGreal proposes that the procedural aspects of evaluation that lead to decisions about teachers must be clearly delineated in any evaluation design.

McGreal is well aware of the pitfalls of a new school district adopting wholesale his system of evaluation. He therefore offers options in various broad areas of teacher evaluation characteristics, or commonalities. If some of these commonalities are construed as a framework for guidance, then the school district may wish to alter its approach to teacher evaluation by choosing among the various alternatives that the commonalities offer. Chapter 4.2 gives a brief account of the eight

commonalities that McGreal has developed for teacher evaluation. These are contained in his 1983 book, *Successful Teacher Evaluation*, published by the Association for Supervision and Curriculum Development (ASCD). The commonalities are based on the belief that teacher evaluation can be both a positive and a productive process. The commonalities are listed below:

1. An appropriate attitude (this makes a clear distinction between formative and summative evaluation by emphasizing teacher improvement rather than a meeting of organizational ends.)
2. Complementary procedures, purposes, and instrumentation (to the extent that procedures and instrumentation fail to fall in line with policy statements, positive attitudes as well as the process as a whole diminishes. Five models for teacher evaluation are cited and briefly examined for their usefulness and adaptability.)
3. Separation of administrative and supervisory behavior (if the outcome of teacher evaluations are to be positive, procedures and instruments must be established that allow the teacher and supervisor to escape the worst effects of a poor, administratively-oriented framework; however, McGreal stresses that dereliction of duty and inability to meet a minimum performance standard must be dealt with administratively.)
4. Goal setting: the major activity of evaluation (as a formal procedure, the goal-setting process is a cooperative activity between supervisor and teacher; various goal-setting approaches are suggested.)
5. Narrowed focus on teaching (any teacher evaluation system that a school or district develops must center squarely on teaching itself to be effective, and ways of doing this are given.)
6. Improved classroom observation skills (classroom observation and professional judgment form the most practical procedure for collecting formal information about teacher performance; four tenets for classroom observation are given.)
7. Use of additional sources of data (these include self-evaluation; peer, parent, and student evaluations; and an artifact collection.)
8. A training program complementary to the evaluation system (the evaluation system is effective only if all those who are to be involved are adequately trained so that steps may be taken to develop and implement a new evaluation system or to revise the extant one.)

As an appendix to his book, McGreal gives an example of an evaluation system (or model) that reflects these eight commonalities. It is a necessary, and valuable, adjunct. Its main purpose is to focus formative evaluation on the delivery system

of instruction, with the staff member and supervisor working together to increase teaching effectiveness and student learning.

Chapter 4.3 Edward Iwanicki: Contract Plans—A Professional Growth-Oriented Approach to Evaluating Teacher Performance Edward Iwanicki advocates using contract plans to guide evaluations of teacher performance. The main point is to provide teachers with feedback they can use to improve their teaching skills and practices. The approach is not keyed to servicing personnel decisions, but rather to fostering professional development.

The focus of the evaluation process is the teaching improvement plan. The plan is keyed to areas of teaching where the teacher needs to improve. The evaluation then assesses and provides feedback on both the implementation of the improvement plan and the impacts on teaching performance and student achievement.

In working with the teacher, the evaluator may use a clinical supervision framework or a management by objectives scheme. The approach is nonthreatening, constructive, and welcomed by teachers. Some writers criticize the approach for its lack of attention to evaluations that can lead to termination of persistently ineffective teachers. Another criticism is that the approach tends to concentrate on teaching styles rather than teaching responsibilities.

Chapter 4.4 Getting Value from Teacher Self-Evaluation by Graeme Withers Graeme Withers of the Australian Council for Educational Research emphasizes that evaluation has importance in the daily lives of teachers. He argues that self-appraisal can and should be held to rigorous standards of teaching performance and student progress and need not be self-serving. He says that such self-appraisal should be ongoing and should provide the basis for planning annual teaching programs based on what worked best in the past. Withers broadens self-evaluation to "co-professional evaluation," evaluations by colleagues of each other's work and against criteria of sound teaching and student progress. He also says that effective self-appraisal and appraisal by co-professionals could provide a basis for holding off external, mechanistic evaluations of teachers by demonstrating that the profession appraises and evaluates its performance from within.

Chapter 4.4 focuses on ways and means of evaluating a teaching program, as distinct from making assessments or measurements of student achievement; however, Withers considers that experience and good practice "in the latter will obviously contribute to the former." This embryonic model attempts to demonstrate a belief that evaluations conducted using internal evidence (from the person being evaluated) and an external view (from a co-professional referee) are potentially more valuable than those carried out by one party only.

Withers is clear that the role of evaluation that he depicts is formative, as it attempts to promote learning and raise professional expertise simultaneously.

Chapter 4.5: Richard Manatt: Teacher Performance Evaluation Richard Manatt, Professor of Education and Director of the School Improvement Model (SIM) for the Research Institute for Studies in Education, Iowa State University, has addressed the growing concern of school districts and the public generally for the need to improve teacher performance. During the late 1970s he accepted and developed the Teacher Performance Evaluation (TPE) approach as a model for teacher evaluation and development. He considered TPE to have a sound theoretical and philosophical base. To promote the concept of TPE, during the 1980's he developed videotapes and accompanying materials for use during seminars and workshops. These activities have resulted in large numbers of administrators and senior educational personnel being strongly influenced by Manatt's cogent approach to teacher evaluation.

The School Improvement Model Project, a very significant undertaking involving two school districts and one independent school district in Minnesota and one school district in Iowa, investigated the effects of a systemwide (or schoolwide) articulated system of administrator and teacher performance appraisal on student achievement. The very real benefits of the outcomes of this study have become important components in national school/teacher effectiveness workshops organized by Manatt and a co-director of Iowa State University's SIM projects, Dr. Shirley Stow.

Although Chapter 4.5 focuses on TPE, the complete picture of Manatt's contribution to the practice of teacher evaluation demands reference to the SIM Project.

Teacher performance evaluation is based upon an analysis of measurement of progress made toward the accomplishment of predetermined objectives or, as Manatt calls them, job targets. This is based upon a process that depends strongly for success on an understanding by both teacher and evaluator of what constitutes effective classroom instruction. It also insists upon effective and efficient use of time. In a Leader's Guide accompanying a videotape for staff development, Manatt (1981, p. 3) stated that to be successful TPE requires

1. Rating scales with criteria based on effective teaching research
2. Lesson analysis in conjunction with skillful observation
3. Coaching and counseling techniques that motivate teachers to change
4. Provision for procedural and substantive due process of law to provide protection for both teachers and educators

Although Manatt's TPE Model has both formative and summative aspects as part of the process, the latter is viewed more as a mechanism for improvement than as an instrument to dismiss poor teachers. This aspect of the process is examined in some detail during the latter part of Chapter 4.5 where the process leading to a summative report about the teacher is given.

Manatt draws a clear distinction between TPE and clinical supervision. The significant difference between the two processes is that teacher performance evaluation is based on analysis and measurement of the progress teachers make toward the accomplishment of predetermined objectives according to policies formulated by the school or school district. Clinical supervision is based on teacher instructional improvement by a professional monitoring process. Perhaps Manatt's most important contribution is that he has placed TPE within the complete context of the school district, linking teacher performance to administrator performance, student achievement and staff development. A well-planned TPE approach where there is thorough commitment by all concerned, helps to ensure that teacher evaluation is both acceptable and rewarding.

Chapter 4.6 Toledo School District: Intern and Intervention Programs

Against all odds and out of a bitter conflict between the teachers' union and school district authorities in Toledo during the 1970s, a spirit of cooperation was born, which resulted in shared decision making in many areas. One such area was teacher evaluation, where the teachers' organization assumed leadership, thus becoming the arbiter both of definitions of teacher competency and of professional standards.

In the Toledo model of teacher evaluation, interest is focused mainly on beginning teachers and those whose performance is below required standards. Skilled, experienced teachers serve as evaluators, and they are trained to a high level of competency and acceptability.

The stated aims of the Toledo model of evaluation is to enhance teacher development. Emphasis on counseling for both probationary and intervention program teachers gives a formative dimension to this model. However, it clearly also serves the purpose of making decisions about a teacher's future. For instance, a teacher will be granted a contract after the probationary internship here only if the evaluation is favorable. Moreover, if a teacher assigned to the intervention program does not receive a satisfactory evaluation, dismissal will follow. The program, therefore, has a strongly **summative** dimension, and it follows that **accountability** is an important outcome.

Both the intern and the intervention programs are well organized and successful. Although there was some apprehension initially about the intervention program, particularly as it could lead to dismissal, the thorough and professional way in which the process was conducted has been reassuring for teachers. Moreover, there is a high level of assistance offered to teachers in the intervention program, which perhaps has been the greatest source of reassurance.

The program has assisted principals in two ways. First, the problem of the poor teacher unwilling or unable to improve his or her performance has been satisfactorily addressed. Second, the difficult tasks of supervising, attempting to improve,

evaluating, and possibly recommending dismissal have been removed from the principal's shoulders.

The chapter concludes with details of a critical analysis carried out by Darling-Hammond et al. in 1984 that found that the validity, reliability, and utility of both the intern and intervention programs were at least satisfactory and generally high. Continued improvements since then have further strengthened critical factors of the Toledo model. Participative decision making has tended to overcome problems before they assume too large a dimension.

Chapter 4.7: Anthony Shinkfield: Principal and Peer Evaluation of Teachers for Professional Development

Chapter 4.7 describes the evaluation model employed over the past decade by a K-12 private boys school in Australia and in many other schools. This model employs principal, peer, and self-evaluation and is focused on professional development. The model's guiding principles include acceptance of the model by school personnel, a constructive orientation, systematic training of evaluators, collaboration and mutual respect between evaluator and evaluatee, clear job assignments and school mission, and confidentiality of the process. The evaluation of each teacher is in-depth, formative, and extends throughout the year. Further, each evaluation is conducted by an Assessment Committee including the teacher, a peer of the teacher's choice, and the principal or other administrator.

It was found that one school administrator could feasibly participate in the evaluations of a maximum of three teachers each year. In many schools implementation of this model would require that teacher evaluations be divided among more than one school administrator. Also, because of the labor intensity of this model, it is sometimes necessary to concentrate evaluation efforts on beginning teachers and those teachers with apparent teaching difficulties, primarily for professional development, but also for decisions about continued employment.

The steps in the model used at the private school include (1) clarification of evaluation policies; (2) initial conferences with teachers to develop a positive climate; (3) a meeting between the administrator and each teacher to select the third member of the Assessment Committee, establishing the constructive orientation of the evaluation for both teacher and school and outlining the procedures to be followed; (4) a meeting involving all members of the Committee to review the proceedings of the first meeting, emphasize the importance of self-appraisal, and lay the groundwork for listing the teacher's major strengths and weaknesses (to be guided by the school's list of important teacher competencies and the duties previously assigned to the particular teacher); (5) about two weeks later, a meeting to review, discuss, and merge the three lists of strengths and weaknesses; closely define and illustrate the important weaknesses; and develop written expectations for improvement; (6) scheduled (approximately monthly) observations of the

teacher's classroom performance and postobservation write-up of the observations, by the administrator and peer; (7) an after-school follow-up conference on the same day as the first observation; (8) an immediate follow-up conference of the second observation aimed at highlighting strengths, reinforcing improvement, and updating competency objectives; (9) subsequent monthly observations and conferences; (10) a wind-up conference to present the final evaluation report, discuss the findings, and determine what further evaluation process may be needed.

The model emphasizes respect for the competence and professionalism of the teacher and a collegial approach to evaluation for professional development. Its orientation is formative as it assumes that the school performed a rigorous job of choosing teachers and generally screened the incompetents out during the selection process. The author observes that teachers who persistently perform poorly can be counseled out of teaching by using the model summatively. However, he recommends that, for schools where such results and needed actions are prevalent, this approach should be supplemented with another more summatively oriented evaluation model.

Chapter 4.8 The National Board for Professional Teaching Standards: Assessing Accomplished Teaching Chapter 4.8 is an overview of the work being undertaken to give recognition to experienced and skilled teachers by the National Board for Professional Teaching Standards. Now the spotlight turns from using standards to assess the worth and merit of systems of teacher evaluation to using standards to gauge the capabilities of teachers themselves.

The National Board for Professional Teaching Standards, which commenced its work in 1988, was a direct outcome of the fears and concerns expressed in two national reports referred to earlier in this Preamble—*A Nation at Risk* (1983) and *Carnegie Task Force on Teaching as a Profession* (1986). The Board has set itself the task of developing approximately 30 assessment packages, all of which provide "high and rigorous standards." The purpose of these assessment packages, which generally depict subject fields appropriate to various levels of student development, is to form a strongly supported basis for identifying successful teachers nationwide.

The ultimate aim of the Board is to influence and improve student learning, schools, school districts, and education (including teacher education institutions) to benefit the quality of life in the U.S. The first, and major, thrust is in the certification of experienced, successful, Board-examined teachers.

For all its bold and commendable aims and activities, the Board is facing some significant problems. The full context in which standards are developed and used must be realized; otherwise, as the Board is discovering, criticisms based on the invalidity of outcomes can arise.

The chapter concludes with a summary discussion of the Board's standards and their validity, together with our thoughts about the benefits and costs to school districts, schools, teachers, and students, of certifying accomplished teaching.

Chapter 4.9 The Tennessee Value-Added Assessment System (TVAAS)—Mixed Model Methodology in Educational Assessment by William Sanders and Sandra Horn This chapter reflects the growing interest and practice in the use of student performance outcomes as one basis for school and teacher evaluation. For example, in Dallas, Texas, under the guidance of CREATE and a National Advisory Panel member, Dr. William Webster, a wide range of student performance outcomes, including test results, forms part of teacher evaluation (using the school as a unit, and not the individual teacher).

In chapter 4.9 William L. Sanders and Sandra P. Horn discuss the background, function, and efficacy of the Tennessee Value-Added Assessment System (TVAAS). TVAAS is a method of assessing the influence of "educational systems, schools, and teachers on the gains their students make on norm-referenced achievement tests." By using mixed-model statistical methodology on collected and aggregated data on teachers and students over several years, TVAAS "can provide measures of the influence of school systems, schools, and teachers on student academic progress."

Although various findings indicated the utility of the Sanders model (as this process has been labeled in Tennessee), until 1988 it was known only to a small circle of educators and some statisticians. In that year, educational reform in the state of Tennessee took a different direction. The Tennessee State Board of Education had published its Master Plan for Tennessee Schools, and the Tennessee Higher Education Commission developed *Tennessee Challenge 2000* for postsecondary educational institutions. The goals and objectives of these governing bodies were coordinated to form an educational framework to address learner needs and expectations from preschool through adulthood. At every level, the need for accountability and assessment was recognized as a central component of educational improvement. Since the focus of the accountability movement was on the **product** of the educational experience rather than the process by which it was achieved, the outcomes-based assessment system developed earlier by Sanders and McLean was closely considered; and in 1991 when the Education Improvement Act was adopted, the TVAAS formed an integral part of the legislation.

Sanders and Horn cite evidence that TVAAS overcomes the major problems that traditionally have been associated with using student achievement data in educational assessment. Use of this model requires huge data sets covering, for example, achievement test results of all students in a state over multiple years plus tremendous computer power. Its feasibility is thus limited to state education departments and large school districts.

Chapter 4.10 An Accountability System Featuring Both "Value-Added" and Product Measures of Schooling, by William Webster and Robert Mendro

Like the Tennessee Value-Added Assessment System, the Dallas Independent School District (DISD) accountability system requires considerable resources to make it effective, and thus is feasible only for large school districts or state departments. However, unlike TVAAS, the Dallas model is not entirely centrally based, since it places strong emphasis on devolution of responsibility for evaluation processes to the schools and school districts. Both models use evaluation processes to achieve accountability in aspects of the educational system.

The DISD model for evaluation began in 1991 with a plan for demonstrable school improvement based on accountability. This is being implemented through a three-tier accountability system. District goals and desired outcomes are established through a districtwide planning process and operationalized through the District Improvement Plan. Each school's role in helping the district to meet its goals is determined through a School Community Council, which ensures involvement at the local campus level. Accountability is operationalized in a criterion-referenced manner through an analysis of absolute outcomes relative to school and district performance on goals specified in both the District Improvement Plan and School Improvement Plans and in a norm-referenced manner through school effectiveness indices. Schools and their staffs are eligible for financial awards based on school performance on the effective indices.

One objective is to identify effective schools and to discover reasons for their success. However, the model has several other useful advantages. One important advantage is that the scheme is designed to foster teamwork among the staff members within a given school; and in order to achieve the necessary improvements in student outcomes, school staff must work together in a coordinated effort. With the school rather than the teacher as the unit, the program does not reward individual competition among teachers within schools. The program also focuses attention on the important outcomes of schooling. The Accountability Task Force, as well as other groups associated with the schools, is given the opportunity to share its views about the purposes and importance of schooling, often based upon weighting the outcome variables, a process that is undertaken annually. It is essential to provide teachers with the information necessary to improve instruction, for it is clear that accountability alone will not improve schools.

Another perceived advantage of the model is that emphasis is given to the effectiveness of schools independently of the status of their student population on the achievement continuum. The techniques reward those schools that impact the most students the most positively. The addition of effectiveness indices thus makes the accountability system valid and fair; each school's performance is judged by comparing its student outcome levels with empirically determined expectations based on *individual* student histories.

References

- Andrews, H. A. (1985). *Evaluating for excellence*. Stillwater, OK: New Forms Press, Inc.
- Joint Committee on Standards for Educational Evaluation. (1988). *The personnel evaluation standards*. Newbury Park, CA: Sage.
- Madaus, G. F., Scriven, M., & Stufflebeam, D. L. (1983). *Evaluation models*. Boston: Kluwer-Nijhoff Publishing.
- Scriven, M. (1993). *Using the duties-based approach to teacher evaluation*. Paper presented at the annual meeting of the Center for Research on Educational Accountability and Teacher Evaluation/Phi Delta Kappa National Evaluation Institute, Kalamazoo, MI.
- TEMP Memo 2. (September 1991). The Center for Research on Educational Accountability and Teacher Evaluation (CREATE). Kalamazoo, MI: Western Michigan University.

Madeline Hunter: Instructional Effectiveness Through Clinical Supervision

As Principal of the University Elementary School, University of California, Los Angeles, Dr. Madeline Hunter gained an international reputation for promoting studies in the cause-effect relationships between teaching and learning. Her work and writings and leadership in extensive inservice and workshop sessions with educators have emphasized the importance of the teacher as an instructional decision maker and have helped to clarify the artistry of teaching.

The "Hunter Models" had their origin during the early 1970s in the Teacher Appraisal Instrument (TAI), which developed into the Teacher Appraisal Instructional Improvement Instrument (TA Triple I). Through formative evaluation techniques, both instruments focus on observing teachers as they give instruction to particular students in particular situations. The TAI and the TA Triple I in some ways became the antecedents of the Hunter Model, which has gone by such various names as A Clinical Theory of Instruction, Instructional Theory Into Practice (ITIP), Mastery Teaching, Program for Effective Teaching (PET), Elements of Effective Teaching (EET), Target Teaching, and the UCLA Model.

Whatever names the model assumes, it is concerned with what a teacher *does* and not what a teacher *is*. The model has not been created for evaluation purposes, but for increasing teaching excellence. The observer makes decisions about growth-evoking feedback and pinpoints effective decisions, reinforces them and states the principles undergirding them as well as inappropriate teaching decisions, and offers

productive alternatives. Moreover, the observer discusses and advises, but does not admonish. To this extent, at least, the model has strong elements of formative evaluation.

In summary, the model

1. allows teachers to identify professional decisions they must make
2. offers causal relationships (often research-based) to support these decisions
3. encourages teachers to use analyzed instructional information to strengthen proven practices that are successful, to develop additional productive alternatives, or to correct their decisions so that the probability of learning is increased

Introduction

An extrapolation from the Hunter Model (or Models) is the **Lesson Design**. This consists of seven elements for planning effective instruction. It is a deliberate focus on principles of learning so that student learning may benefit and accelerate. The recommendations for Lesson Design that are discussed in this chapter must be placed in their correct context. The Lesson Design is only part of Hunter's complete program, although it must be recognized as one of its most essential and most practical aspects. The evaluation of a teacher's lesson for instructional improvement is based by some users of the Hunter model on concepts and procedures undergirding the Lesson Design. However, Hunter herself never promoted this usage.

Proponents of most of the models discussed in this book emphasize the importance of inservice training for personnel before a model is even trialed in the school or school district. Madeline Hunter stressed this more than most. She insisted that educators who are to become leaders of the program must have intensive inservice training over a lengthy period of time before they attempt to influence classroom teachers. With complete training, leaders will be able to instruct teachers as decision makers in the learning process, work with them in the principles of learning and their implementation, develop and improve observation skills, and have the ability to carry out supervisory conferences for improving teacher excellence at relevant stages in the model's progression.

The Hunter model, then, relies on the strength of inservice training of leaders and others involved in the program. Leaders must learn basic principles of human learning and the various model components that flow from them; namely, teacher decision making, observation skills, and the ability to organize and conduct appropriate kinds of growth-evoking supervisory conferences.

This chapter will discuss the various components of the Hunter Model with particular emphasis on the Lesson Design. Although the various components will be treated separately, it should be understood that once leaders are trained and the model is about to be implemented, the specific needs of teachers and schools may dictate where time and effort are to be spent to make the approach work successfully. It should be noted, moreover, that whereas supervisory conferencing skills are contained in a separate section of the chapter, the various types of conferences that take place between the administrator and the teacher will be determined by a teacher's sophistication in implementing cause-effect relationships in students' learning.

Inservice training: Leaders—Then Teachers

For its success, the Hunter Model requires considerable time, resources, and involvement by all staff, as well as major commitments to the success of the process. It is essential that leaders are fully conversant with the procedures that are to ensue; that teachers, as evaluators of their own instructional programs, must know what is to happen and what is not to happen in the classroom in the way of student learning; that terminology is understood and observation procedures are agreed on; and that sufficient time is set aside for observers to conduct the process thoughtfully. Hunter stated five attributes that are critical to the program as it attempts to increase teaching effectiveness:

1. A specific research-based content that is able to be translated into classroom implementation and then validated by observation of subsequent teaching performance
2. Leadership qualified to teach professional content, monitor progress, and keep the program on track
3. A written plan that details all aspects of the program including a time line with formative evaluation check points
4. An adequate budget so that the time and personnel needed to accomplish the program are available
5. Knowledge of the problems common to such a program so that solutions for those problems become a deliberate part of the plan (1977, p.2)

Obviously, both training of personnel and commitment of valuable resources will be needed if the program is to achieve its stated intentions.

The thorough training of leaders is the starting point.

Preparation of Leaders

If a productive inservice program is to develop, district educators with potential for leadership must be recruited and trained. While the initial impetus may come from a person with expert knowledge of the model, sustained growth will occur only with local commitment.

Hunter suggested that potentially these leaders should progress through seven phases, with proficiency being acknowledged at the conclusion of one phase and before another is attempted.

Phase I. Comprehension of the Inservice Content Participants acquire knowledge and comprehension of the cause-effect relationship of teaching and learning as the basis for artistic teaching. Concepts and generalizations are labeled and explained by participants on the basis of examples that are presented by videotaped observations in teaching episodes.

Phase II. Internalization of Inservice Content In this phase, participants demonstrate the use of the cause-effect relationships in teaching and learning while teaching students in a sequence of consecutive lessons. Participants work on content with which they are familiar; emphasis is therefore on practicing and understanding the skills of effective teaching while working with content that is new to the students (but not to the participating teacher). Participants are observed and subsequent modifications of their teaching performances are made as a result of feedback from knowledgeable observers.

Phase III. Comprehension of Observation and Feedback Techniques Here, focus is placed on comprehension of the skills necessary to analyze another's teaching performance. The teacher being observed is offered constructive feedback, which models the same principles of learning that are expected of the teacher. The skills to be learned may be based upon a videotape of a teacher. This allows the leader-in-training to observe teaching episodes, to capture with a script of a tape the sequence of what has occurred in the tape and, from the tape, to label teaching-learning behaviors. Arising from this will be different types of feedback communicated during conferences with the teacher who was observed.

Phase IV. Feedback from Knowledgeable Observers Leadership training also involves participants in teaching lessons and becoming the recipients of feedback during conferences with knowledgeable observers. This allows participants to continue the understanding process that began in Phase II and to experience receiving feedback as it is presented in the observation-conference process of Phase III.

Phase V. Internalization of Observation Feedback Techniques During this phase, participants acquire the knowledge and practice to understand completely the skills that are necessary to conduct a "growth-evoking" instructional conference. The skills to be learned are listed below:

1. observing and recording by script of a tape and analyzing a videotape of a teaching episode
2. designing growth-evoking objective(s) and strategies for achieving appropriate objective(s) in a subsequent instructional conference
3. conducting the conference and modifying strategies as a result of sensitive diagnosis from observations of the teacher's own response(s)
4. subsequently evaluating the success of the conference and generating information and ideas that can be used in subsequent conferences with that teacher, and that can also be used in a generalized sense to increase the success of conferences with other teachers

This developing skills in conducting conferences phase involves the participant in being observed in a conference and then making modifications as a result of feedback from knowledgeable observers. It is a practicum for developing observation and conference skills. Moreover, participants practice observing each other teach and conducting instructional conferences. Eventually this phase should lead to observations and conferences with teachers who are not involved in the leadership training, a process that also receives feedback from knowledgeable observers.

Phase VI. Comprehension of Presentation Skills for Staff Development

This phase focuses on the skills necessary to design and implement a staff development program. It is important that leaders-in-training become familiar with the research base of current professional knowledge to support elements of the Decision Making model and to respond adequately to questions they may encounter in their future leadership role. Organizational abilities also rank high as components of successful presentations to others. Unambiguous examples related to theory and to the participant's personal and teaching experience need to be generated and rehearsed.

Phase VII. Internalization of Presentation Skills for Staff Development

This final phase develops a leader's performance behaviors that model artistic practice of the professional content that will be presented to others in staff development conferences. Behaviors include the development of group dynamic skills, small and large group presentation mastery, leading discussions, and monitoring the quality of learning that will enhance participant achievement.

Hunter pointed out that there is a marked difference in skills required and performance complexity between Phase III (Comprehension of Observation and Feedback Techniques) and Phase VI (Internalization of Presentation Skills for Staff Development). Any attempt to make the quantum leap from an earlier phase to Phase VI without building intervening skills (which are often based on errors that must be corrected) is a recipe for disaster. In fact, Hunter advised that for most educators the progression from Phase I to Phase VI should take a minimum of two years of study, practice, and complete understanding, aided always by continued coaching from knowledgeable observers.

After the completion of initial training, the leader should be in a position to gain the cooperation of others to ensure the implementation of the program into a school district and schools themselves. One vital aspect is the professional preparation of teachers, which will occur at an appropriate stage of planning for implementation of the model.

Preparation of Teachers The most important aspect of the preparation of teachers is the development of research-based skills necessary for them to become the decision makers about the instructional process. This is so important that the next main section deals with this topic.

Leaders must make teachers aware that they can learn the skills to be responsible for student learning. To support or augment preservice instruction, leaders may need to carry out inservice staff development emphasizing the basic skills required for any teaching: diagnosing learners, analyzing the learning task, sequencing learning, using learning principles that affect students' motivation, rate and degree of learning, and so on. These basic approaches to teaching are the foundation of effective and artistic teaching. However, they are not the total of what is known as the cause-effect relationship between teaching and learning; consequently, there must be ongoing staff development to promote continuing professional growth. This will be undertaken in conjunction with other aspects of the Hunter model. Thus, staff inservice must be organized in a clearly defined fashion that emphasizes teacher decision making in the cause-effect relationship of teaching and learning and the translation of such relationships into artistic teaching.

Throughout teacher inservice activities, the leader's performance must model effective teaching to be convincing; the importance of preparation of leaders is therefore once more underlined. The leader must have learned to effectively employ principles of learning for teachers, just as teachers must learn to use those principles for students as they endeavor to learn.

Planning, Implementing and Evaluating the Inservice Program Any successful inservice program that is designed to increase instructional effectiveness of teachers must be grounded in sound planning, implementation, and evaluation.

Leaders from both the administration and teacher organizations work in a collaborative rather than adversary relationship from the inception of planning. An outside consultant, well versed in the principles of inservice training, may help through the early stages of planning and provide remedial feedback to allow implementation of plans to go forward smoothly. Giving such advice to district leaders is not essential, however, if they have undertaken a complete course of training themselves.

Those involved in the program should be willing to be committed for a period up to five years to allow continuity of growth by knowledgeable personnel. Hunter suggested that some, or all, of the following could be involved: leaders from the administration and teacher organizations, central office and school administrators, volunteer administrators and teachers, and future trainers who are selected from the volunteers to develop the knowledge and performance skills necessary for leaders of district staff development.

Much of the planning will revolve around content that is known to be useful in effective and artistic teaching. After introduction as a teacher decision model, the actual order in which content is learned will depend on the needs of the district and the judgment of the trainers. After an introduction to the categories of teaching decisions, district participants will learn (or reinforce their learning) about common categories of effective teaching, such as principles of motivation, elements of planning for effective instruction, extending students' thinking, transfer and retention of knowledge, lesson analysis, and types of instructional conferences. As a district commitment, inservice should take place during the work day and teachers should be accountable for learning and implementing the content.

Sufficient time must be allowed for many systematic follow-up observations of the teachers', administrators', and district leaders' implementation of the inservice content. Feedback, reinforcement, remediation, and change usually follow. Although the time required for observation and feedback is one of the most costly factors of the program, it is essential for its success.

Because the program must be continuous, and developing, it is likely to be expensive in terms of time. Ad hoc diversions and the latest "in-thing" will drain off professional energy. To allow leaders and practitioners time to translate what they know into what they do, the process cannot be rushed. Budgetary support in terms of time and personnel must be adequate for inservice and implementation. There must also be adequate resource provisions made for the formative evaluation of what is occurring during the inservice program. Like the planning and implementation of inservice, the evaluation should be so thorough that the district has a clear indication of the extent to which the continuing inservice courses are successful in classroom implementation. Budgetary considerations in terms of time, personnel, and finance must be made for formative evaluation processes. Time, or lack of it, is often the main stumbling block. Hunter insists, however, that follow-up

of participants' inservice performance (leaders, teachers, and school and district administrators) with reinforcement and/or remediation is often minimized; and yet this is a critical element for the success not only of present and future inservice courses but for the model as a whole.

The Teacher as the Decision Maker

Much of the content of inservice professional development is directed toward the Lesson Design. As has been pointed out, emphasis is given to the content that is known to be useful to support effective and artistic teaching theories. A specific instruction may also be given in the Teaching Appraisal Instrument (TAI) to use observed classroom behavior and data to answer five questions:

On a vertical axis, indicating the "what" of teaching-learning:

1. Is teaching-learning time and energy focused on the intended objective?
2. Is the objective at the appropriate level of difficulty?
3. Is there constant monitoring and adjusting?

On the horizontal axis, which shows the "how" of teaching-learning:

4. Which principles of learning are being used productively?
5. Which principles are being abused or ignored?

Instruction may be given in the Teacher Appraisal Instructional Improvement Instrument (TA Triple I), which is a diagnostic prescriptive and/or evaluative tool. The TA Triple I accommodates a wide range of data collected from many observations of particular students and situations. These data are interpreted in terms of stationary reference points that have been established by research.

One important aspect of the TA Triple I is that it can be used "to improve instruction by helping a teacher know whether the teacher-learner energy is focused on the intended learning or is being dissipated, which learning principles are being used appropriately to further student learning, which additional principles could be used to accelerate that learning, and which principles, if any, are being ignored or abused, thereby interfering with intended learning. An extremely important contribution of this instrument is the articulated information of what a teacher is doing well and why it is successful" (1976, p. 10).

Whether the TAI, the TA Triple I, or other approaches such as ITIP (Instructional Theory Into Practice) are used or emphasized, the fact remains that teaching is decision making, and successful learning results from successful decisions being made by teachers.

Focus on the Teacher There is no doubt that of the host of factors that influence a student's successful learning, the teacher is the most important. What a teacher says and does, and how well he or she says and does them, will determine a student's progress in learning. The prime responsibility is therefore placed on the teacher to make effective decisions, a process that is the very essence of the Hunter Model.

It follows that the program leader, often the school principal, must be fully conversant with the kinds of decisions, and the reasons for these decisions, that the teacher has to make. Instruction in this area would have been given during preceding inservice training. Space will not allow a full explication of the process here. In brief, it is an analysis of decisions made in teaching by the teacher.

To help student learning a teacher must evaluate the instructional process, based on the investment of the learner's time, to determine whether such investment is in keeping with current learning knowledge. Questions like those listed below are asked:

1. Is the instructional process proceeding toward a perceivable objective?
2. Is the instructional objective at the right level of difficulty for the learners who are investing time?
3. Is there constant monitoring of the degree of achievement of the objective so that the instructional process may be accelerated or slowed down?
4. In which ways are the time and energy expended by learner and teacher consonant with principles of efficient and effective learning?
5. If there is dissonance between time and energy expended and principles of learning, which principles are being violated?

Although it is the task of the leader to guide the teacher toward addressing these, and other, questions relevant to teaching and learning, it is the teacher who knows individual students in the classroom situation who must provide the answers. Decisions in teaching are relativistic and situational. The focus is on the teacher in such decision making. Removed from the situation, the most eminent learning theorist cannot make a decision as appropriate or relevant as the sophisticated teacher on site who has both the information and the skills necessary to make decisions with high probability of productive outcomes.

Training for Decision Making During the late 1970s, the approach frequently used for teaching appraisal for instructional improvement was the TA Triple I. Into the '80s and '90s, the ITIP became more popular in some school districts. They both serve a similar function of developing effective teaching through task analysis of the complexity of learning and then diagnosing to identify which components a student has achieved and which remain to be accomplished. Films and videotapes make it possible for a teacher to see professional decision making implemented in

a typical classroom. Such a process helps the teacher to improve teaching skills, with guidance provided by a mentor.

Using one of the instruments mentioned, the trained observer can identify teaching behaviors that research and classroom evidence would support as increasing the chances of learning. The important thing is that these behaviors must first be fully understood by the teacher who, having chosen those that are most effective for student learning, increases the deliberate and appropriate use of these principles and approaches in the future. Moreover, the TA Triple I, in particular, will reveal teaching decisions and actions that, although often unintentional, interfere or hinder a student's successful learning accomplishments.

Hunter sums up teacher decision making and its associated training by stating the four essential components leading to this aspect of professional development:

1. Identification of the decisions a teacher must make
2. Inservice that enables the teacher to combine science and art in teaching
3. Films and tapes that provide opportunities to predictably "see" how it looks in the classroom
4. A diagnostic-prescriptive instrument that provides knowledge of results in professional performance (1979, p. 67).

Lesson Design

The essence of Lesson Design is that teachers learn to spend instructional time in areas where there is reasonable support for lesson plans having a direct impact on student learning. Although the focus is placed on the teacher and the teacher's decision making, the approach provides an opportunity for leaders (whether the principal, supervisor, or trained colleague) to work closely with the teacher to achieve planned ends.

Based on thoughtful planning, the Lesson Design model is both practical and efficient and is applicable to all modes of teaching. The seven elements of the model reflect practical characteristics that have made effective research acceptable to practitioners. Learning theory perspectives predominate. Because they make sense and parallel accepted practice in schools, their credibility has been very strong.

While a successful lesson may be planned and followed in the classroom by incorporating the elements, in reality they form an appropriate framework for planning virtually any kind of lesson (discovery, teacher directed, cooperative learning, etc.) at any grade level and in any subject area. In other words, the seven elements build a teaching focus containing essential teaching skills that are applicable in any situation.

It is assumed that before a teacher begins to plan a particular lesson, the primary objectives of that lesson will already have been determined. With that achieved, the following elements are used to design a lesson that is considered most effective to meet the planned objectives. These elements are described separately to determine whether or not they are appropriate for the objective (or objectives), bearing in mind the particular characteristics of the students to be taught. Any element may be included or excluded, but it must be integrated into the artistic flow of the lesson as a whole.

Element 1. Anticipatory Set During the first minutes of the lesson, students must be mentally prepared to learn and immediately encouraged to concentrate on what is to follow. Effective activities to develop an anticipatory set will include focusing the students' attention on the ensuing learning, possibly providing a very brief practice on what was previously achieved (or on related learnings) and in other ways developing a readiness for the instruction to follow. It is important that students know the relevance of what they are to learn and also that they gain a sense of continuity. There must be a relationship between today and yesterday if yesterday's learning is to facilitate today's. At times yesterday's learning can impede today's, and so it is not referenced.

To help continuity, one effective technique that has emerged from literature and research is the teacher's use of statements that provide important cues for students.

Element 2. The Objective and Its Purpose An anticipatory set of statements leads into the lesson's objective and its purpose. Information must be placed into perspective and its value perceived.

If appropriate, this element requires the teacher to communicate to students what they will learn by the end of the instructional period, why the accomplishment is an important and useful development, and how it related to their lives. In most cases, students have both a right and need to know how a present lesson relates to past instruction and why it is important to them at present and in the future.

Element 3. Instructional Input During the planning of this step, the teacher must determine what information, skills, or processes the student requires so that the present objective may be achieved. Information given to students may be based on what they should already possess, or what is presented may be a new experience for them. Students will find it difficult to achieve an objective without having been taught the prerequisite background information. In some ways Element 3 is an explanation-demonstration stage in teaching. It can also occur through discovery or learning.

Once the necessary information has been identified, the teacher selects methods to accelerate and to check student understanding and learning. The possibilities are

endless apart from verbal communication: books, films, records, diagrams, and artifacts.

Element 4. Modeling As a strong grappling hook of learning, the teacher supports examples with the perceptual input of modeling. It is most helpful for students not only to learn about something but also to see or hear examples of an acceptable finished product or process, be it a story, model, diagram, picture, or scientific experiment. It is equally important that they perceive a process in action, such as articulated thinking during the process of an assignment, e.g., how a goal is thrown in basketball or how a graph evolves.

Modeling should be accompanied by verbal input, such as labeling the critical elements of what is occurring, so that students are able to focus on the essential aspects of the lesson objective throughout, rather than on nonrelevant factors in the process or product.

Element 5. Checking for Understanding So that appropriate instructional decisions are being made, the teacher needs continuously to monitor students' level of comprehension. Since it is likely students learn best when they are first introduced to new material, if they do not understand what has been presented, it is best for the teacher to reteach the material immediately.

This element requires the teacher to check by objective evidence the students' grasp of essential information and also to observe their initial performance to ensure that they show the skills necessary to achieve the instructional objective. For example, the teacher may sample knowledge by posing appropriate questions, have students signal responses, or elicit individual private responses. It is especially important that nonrespondents to general questioning are included in signaling or seeking private responses. Signals or brief written responses, which the teacher can quickly peruse, can indicate the extent of student learning.

Element 6. Guided Practice When the teacher perceives that a satisfactory operational level of understanding has been reached or appears to be attainable, it is essential that students be given the opportunity to practice the new skill or its application under teacher supervision. This guided practice, or controlled practice, helps substantiate or correct students' initial attempts in new learning.

The teacher elicits group practice by moving among the students, checking individually to see whether the new instruction has been understood before allowing them to practice independently. Students then perform sufficient further examples so that clarification or mediation may occur immediately. In this fashion, the teacher is assured that students are able to perform the task satisfactorily without assistance and that they will not practice mistakes when working by themselves.

The teacher works with the students providing support, encouragement, individual assistance, or further teaching as required.

Element 7. Independent Practice A student who is able to perform without significant errors, confusion, or embarrassment is ready to develop fluency and artistry while practicing without the help of the teacher. Independence is the true hallmark of effective learning. An independent student can be given an assignment to develop fluency with the new skill or process without the direction of the teacher. As a practical example, students should never be sent away with homework containing tasks that they have not demonstrably understood in class.

Summary Remarks Although Hunter developed and labeled the Teacher Decision-Making Model, effectiveness research discussed in Chapter 4.2 by McGreal has more recently provided the hard data. The sequence of decisions for an effective lesson presented here is basic to sound learning. Teachers need to evaluate their lesson plans and develop procedures for constant instructional improvement.

Although thorough planning may have overtones of sound professional workmanship rather than naturalness, artistry in teaching is impossible without a thorough instructional design. Hunter often stated that both the science and the art of teaching are essential and that the seven elements mentioned in this section, which promote effective instruction, constitute the launching pad for creative student attainment.

It is worth repeating that a teacher does not have to include all seven elements within a single lesson. In some instances, lessons will incorporate only the first three elements, or only guided practice, although it is anticipated that over a series of lessons, as students progress toward achievement of complex learning, all seven elements will be addressed. This point leads us to a consideration of the very special tasks of observers in the total Hunter program of teacher assessment.

Observation of Teachers

Leaders of the staff development and evaluation program must be sophisticated, perceptive, and sensitive observers. Unless they possess these essential characteristics, it is doubtful whether the program as a whole will have a chance for success. It is essential that extensive inservice training has sharpened or developed leaders' observation and coaching skills.

It has been sufficiently emphasized that the teacher is responsible for instructional decisions. Observations by a knowledgeable observer and the conferences that follow enable the enhancement of collegiality and continuing growth essential to the professionalism of teaching.

Planning Conference A planning conference is one in which the observer and the teacher collaborate in the design of a subsequent lesson for successful learning outcomes. Although the teacher is responsible for initiating decisions related to cause-effect relationships in learning, responsibility for successful learning outcomes from a planning conference is, in many ways, the joint responsibility of both the observer and the teacher.

Planning conferences are an excellent opportunity for the observer to renew teaching skills and for the teacher to seek guidance from another professional person and to experience the stimulation of collegiality.

A planning conference will once again stress the importance of cause-effect relationships in teaching and learning. To this end, films of teaching can be analyzed and productive teaching behaviors identified and labeled. Inservice or staff meetings centered on such aspects of teaching can become effective introductions to the observations that follow.

Hunter drew a sharp demarcation between a planning conference and a preobservation conference. She saw a preobservation conference as unnecessary. In fact, she stated in clear terms that the preobservation conference should be eliminated as it is unnecessarily time consuming, builds bias in both teacher and observer, lowers the level of trust and rapport, and is likely to defeat the purpose of the approach, which is to constantly promote escalating instructional effectiveness. Moreover, an observer's stance must always be analytical, not critical.

Aspects of Observation of Teaching Depending upon the number of trained observers in a school or the amount of time available to trained observers such as principals, all teachers should be observed and assessed many times for modification, reinforcement, or enhancement of effectiveness of instructional methods. Included in classroom observations will be excellent teachers, since they not only provide fine examples for future modeling, but also need to continue to grow and not regress from their peak of perfection, something that may occur without ongoing analysis of their teaching.

Research and study have shown that there are several good qualities that best teachers have and that will be contained in any strong learning situation. Hunter identified five qualities of good teaching (TAIII):

1. *Teaching to an Objective*

When a teacher is teaching to an objective, each component of the lesson leads to the next and all are connected to a major objective or related objectives. With adherence to time such an essential factor of a successful lesson, the teaching to specific objectives is very important. It is not difficult for an observer to determine whether the lesson objective is being followed

and logically developed or whether time is being frittered away by off-target tangents.

2. *The Objective is at the Correct Level of Difficulty*

Discussion may well have occurred during the planning conference about the appropriate level of difficulty of content and language in respect to a particular group of students. Along with the observer, the teacher should discern whether students have grasped content or concept and whether they are ready to move on to the next logical step of the lesson. The observer must ask the question: "Are students challenged and interested?" This is only one of many relevant questions.

3. *Is There Monitoring and Adjustment?*

This quality of good teaching follows from the previous one. The good teacher will monitor and adjust so that the level of lesson continues to be appropriate to student abilities and readiness for new information. A teacher's training for on-the-spot decision making, as outlined in the previous section, is essential in this respect.

4. *The Teacher Applies the Principles of Learning*

A good teacher will apply the critical principles of learning such as motivation, rate and degree, retention, and transfer. Both the observer and teacher must be aware that these essential principles, which are the bases of sound learning, clearly are evidenced, when appropriate, during the course of the lesson.

5. *The Teacher Continues Professional Growth*

A most significant aspect of observing is helping teachers to add new skills to their reporting of professional alternatives and to see things they may not have seen and which they can correct. A professional person is one who retains and implements an interest in improvement and willingly seeks the analytical advice of a perceptive observer.

Observation as Interpretation Any astute observer should interpret each part of a lesson in relation to preceding and subsequent parts, and each behavior in terms of prior and subsequent behaviors. It follows that while the teacher and observer may be interested in the development or refinement of a particular instructional skill or technique, the observer's focus must also include all *other* aspects of a teacher's performance. In other words, viewed in isolation, no teaching skill can be interpreted accurately. Consequently, checklists in observation are inadequate. While there may be agreements reached in advance for the observer to focus on one skill or technique in depth, such an observation cannot be undertaken in isolation from all else that occurs during the course of a lesson.

During this part of the Hunter Model the observer becomes an appraiser who endeavors to see whether the teacher is demonstrating competence in particular

skills. To the extent that evaluation is taking place, the emphasis is always on positive outcomes that Hunter often interpreted as perceptive and useful feedback to improve both teacher skills and learner potential. Observation, then, is formative evaluation. The more closely and perceptively that an observer interprets what happens in a classroom according to its effect on student learning, the more the teacher will gain from the conference that follows the lesson.

As the observer watches and analyzes teaching, a tape must be transcribed to capture information to be used during the teacher conference. An observer's memory is often inaccurate. The observer will also record answers to questions such as these:

- If appropriate, what is the main objective of the lesson?
- Has time been spent only on the target objective and complementary issues?
- Have student responses indicated a successful achievement of the target objective?
- Are there strong indications that most (if not all) students understood the target objective?
- Has the teacher (by questioning or other methods) validated student achievement and has the level of difficulty been adjusted if appropriate?
- Are the principles of learning being utilized?

The tape transcription will supply evidence as to the extent to which objectives have been met and other specific observations that can be used during the teacher conference that follows. A summary form may be used if a record is needed.

Summary Sophisticated observation and feedback are the essence of clinical supervision. In terms of clinical practice, the interaction of professionals, based on an understanding of particular students and a specific content, promotes continuing professional growth and student learning. Observations based on specific cause-effect relationships in teaching and learning are the antithesis of the vague generalizations and admonitions that have tended to dominate the supervision of teachers. The observation is a commonly shared experience, and the analysis and assessment of what happens during a classroom period accelerates teaching excellence if it is carried out well.

Types of Supervisor-Instructional Conferences

The most important purpose of a supervisory conference is to promote the teacher's growth in effective instruction. These conferences are both diagnostic and prescrip-

tive with the baseline being the enhancement of the quality of education in the school.

All supervisory-instructional conferences should have a primary objective. This does not mean to say that there will not be supplementary objectives; but no objectives divergent from the primary purpose should be included.

There is always the assumption that teaching is a performance behavior and can best be improved through the analysis of that behavior. In order to secure the information needed for a successful conference, the supervisor must first observe and script tape at least one episode of teaching. In Hunter's experience, 10 to 20 minutes of observation yield at least an hour of conference material. With a developed ability to analyze and observe an episode of instruction, the observer uses diagnostic judgment to select which of 5 possible communications should be the main purpose of the instructional conference. While these objectives are not mutually exclusive, they each generate different potential learnings for the teacher.

There are five types of Instructional Conference possible. All conferences are focused on transfer to future teaching.

- "A" Conference: the observer reviews with the teacher the successes of the lesson in terms of teacher behavior, which leads to student behavior; labels the behaviors; and gives the generalization that undergirds that behavior.
- "B" Conference: the observer helps the teacher develop a wider range of effective teaching behaviors.
- "C" Conference: the teacher identifies satisfactions with the lesson and plans strategies in collaboration with the observer to eliminate dissatisfactions.
- "D" Conference: The observer identifies and questions behaviors and, if appropriate, suggests alternatives for less effective aspects of the lesson that may not have been apparent to the teacher.
- "E" Conference: the observer reviews the lesson of an excellent teacher who then selects the next steps for expanding professional growth.

1. Type A Instructional Conference Communication During this communication, the observer identifies, labels, and explains the teacher's effective instructional behaviors so that these successful techniques are deliberately and appropriately applied by the teacher in future lessons. To achieve this objective, the observer

focuses on those aspects of instruction that were effective, explains why they worked well, and identifies future conditions where they could be effective.

The observer cites examples of good teaching from the transcription of the tape, such as ways in which the teacher recognized individual children's talents, reinforced a correct response from one student for the benefit of all, questioned techniques that elicited responses indicating students' failure to grasp a new concept, and appropriate timing to achieve the lesson's objective. Hunter pointed out that for a first conference, or with apprehensive or defensive teachers, Type A conference communications may be the sole outcome of what should prove to be a productive instructional experience.

2. Type B Instructional Conference Communication Here, the observer endeavors to stimulate a range of alternative, effective teaching responses. If the teacher and observer work together to generate instructional techniques in addition to those that were effective in the observed lesson, the teacher will have a wider scope of instructional approaches to use in future lessons.

All of us have a tendency to become habitual in approaches and responses. If teachers are able to become more flexible in their patterns of presentation, newly acquired creativity and flexibility can enhance student learning.

It should be noted that Type A and B conferences focus only on the teacher's effective instruction. Thus, Type B conference communication must be construed by the teacher as a positive instructional stimulant.

3. Type C Instructional Conference Communication This conference aims at encouraging teachers to identify where they were satisfied or dissatisfied with a lesson so that, in collaboration with the observer, strategies may be developed for enhancing successful, or reducing or eliminating future unsatisfactory outcomes.

The parts of the lesson discussed are initiated by the teacher who states what went well or did not go as well as anticipated, and teacher and observer generate what could be done for improvement. If this type of conference is to succeed, a basic requirement is that both the observer and teacher comfortably discuss the situation. While the teacher initiates the topic in the discussion, the observer offers guidance.

4. Type D Instructional Conference Communication Emphasis in this conference is placed on identifying any questioned aspects of the lesson that may not be evident to the teacher, checking the reasons for the teacher's decisions (which may be found to be inappropriate) and developing alternative approaches that have the potential for increased success.

The observer may not wish to suggest a range of alternative teaching behaviors from which the teacher can select substitutes for those that failed to succeed. Again,

it should be emphasized that it is the teacher who is the decision maker. If the teacher believes that the alternative approaches are more likely to benefit student learning and learns to use these behaviors for future lessons, increased learning of students will undoubtedly occur.

Although the Type D Conference communication is focused on questioned teacher behaviors, it need not be negative. In fact, it can be a relief for a teacher to know that support and advice is being offered and that clarification and suggestions for ineffective aspects of the lesson have been offered. Very often, finding out what caused the trouble is the only information necessary to eliminate it.

The sensitive leadership of the observer is essential during this type of communication. In all probability, the observer has the sole responsibility for identifying cause-effect relationships between teaching and student responses. It is also likely that the observer will have the task of generating alternative teaching decisions and behaviors and helping the teacher conclude that these may be more productive.

Perhaps the observer's most difficult task is to direct advice for change that is based on that teacher's particular style and techniques and *not* on how the observer would have taught the lesson (or other lessons).

5. Type E Instructional Conference Communication Type E communication promotes the continuing growth of teachers who are excellent. Gifted teachers, like gifted students, must be encouraged to continue their growth by selecting the next steps in their professional repertoire. This communication is designed to promote growth beyond that which the teacher alone can generate.

This communication should be creative and stimulating. The observer may suggest that the teacher allow a particular lesson to be videotaped for the benefit of others or that student teachers be permitted to observe some lessons.

While it is acknowledged that sometimes it is difficult to identify ways of encouraging the growth of an excellent teacher, both the observer and teacher should accept the challenge and, working collaboratively, develop imaginative ideas that may be put to practical use.

Summary of Instructional Conference Communications The intentions of the five types of communication are not mutually exclusive although one may be the principal objective. Four of the five conferences are positive and the fifth (Type D), if planned and undertaken wisely and professionally by the observer, has the potential to also be positive.

Once again, it is stressed that the observer needs to possess a range of skills so that the expert analysis of the lesson being observed becomes the basis for a productive conference. Skills include the analysis of instruction in terms of cause-effect relationships and, where appropriate, generating solutions to instructional problems. Important, too, are the strength and sensitivity of communication skills.

The Evaluative Conference If the first function of a supervisory conference is to promote the teacher's growth in the ways outlined in Type A, B, C, and E Conference Communications, then a summative function of a year's supervisory conferences, according to Hunter, is evaluative.

The purpose of an evaluative conference is to place the teacher on a continuum from "unsatisfactory" to "outstanding" and to give the teacher the opportunity to examine the evidence used to reach a particular conclusion. An evaluative conference is often not done after an observation, but is the summation of issues arising from a number of instructional conferences. Thus, information given to the teacher, and conclusions reached, should not come as a surprise, since supporting evidence should have been discussed during earlier conferences. Hunter pointed out that the summative evaluative conference may be the culmination of a year's diagnostic and collaborative work in which the observer/supervisor and teacher have shared responsibility for the teacher's continuous professional improvement.

If a teacher's strengths and, if present, instructional deficiencies are seen in a balanced perspective of various types of instructional conferences and are not highlighted for the first time during the summative evaluation, then professional development most likely will occur. A final evaluation where weaknesses rather than strengths are emphasized will not only be negative but also will preclude the chance of professional growth.

Conclusion

A review of the chapter clearly indicates that the Hunter program for clinical supervision is an extensive, productive, but costly (in terms of time, energy, and money) undertaking by a school district. Unless there is a sincere commitment by all administrators, teachers, and the teacher organization, the model can certainly falter, and probably fail. Such factors as the time and financial outlay, the training period essential for leadership proficiency, and the rigorous nature of clinical supervision itself and attendant conferences are undoubtedly daunting.

On the other hand, the benefits are considerable. These include increasing knowledge and perceptions about the nature of the teaching and learning process and its causal effects, a heightened awareness of the range of instructional skills and techniques, the value of administrators and supervisors working collaboratively with teachers, the professional development of teachers, and increased potential for student learning. A further major advantage is that the training period undertaken by educators who are to become knowledgeable observers and assessors develops and sharpens a wide range of skills relevant to clinical supervision as well as knowledge about education generally, and personal communication skills in particular.

Unlike some other teacher evaluation models or teacher improvement approaches that have evaluative components, the Hunter program is designed to make all outcomes as productive as possible. One of the most productive outcomes is the focusing of decision making on the teacher who is encouraged, and even compelled to analyze teaching situations for both learner and teacher enhancement. Through clinical supervision and astute observations, teachers learn to use effective techniques more often and to do so deliberately.

Hunter and her colleagues at UCLA helped to demonstrate that there is a science undergirding the art of teaching, a science that predictably can be acquired. When effective teaching is observed and becomes part of the teacher's repertoire of skills and techniques, artistry in teaching becomes increasingly apparent.

References

- Hunter, M. (1976). Teacher competency: Problem, theory and practice. *The Early and Middle Childhood Years of Schooling*. Journal of the College of Education, 15(2). Columbus, OH: The Ohio State University.
- Hunter, M., & Russell, D. (1977). Planning for effective instruction, (lesson design). *Instructor*. Dansville, NY.
- Hunter, M., & Russell, D. (1977). *Critical attributes of a staff development program to increase instructional effectiveness*. Unpublished occasional paper, University Elementary School, University of California, Los Angeles.
- Hunter, M. (1979). Teaching is decision making. *Educational Leadership*, 37, 408-412.
- Hunter, M. (1980). Six types of supervisory conferences. *Educational Leadership*.
- Hunter, M. (1983). *Mastery teaching*. El Segundo, CA: TIP Publications.
- Hunter, M. (1985). *Prescription for improved instruction*. El Segundo, CA: TIP Publications.
- Hunter, M. (1993). Enhancement of teaching through coaching, supervision, and evaluation. *Evaluation Perspectives*. Center for Research on Educational Accountability and Teacher Evaluation (CREATE), Kalamazoo, MI: The Evaluation Center, Western Michigan University.

Thomas McGreal: Characteristics of Successful Teacher Evaluation

Thomas McGreal has worked with hundreds of school districts over the years to encourage the design and development of realistic and effective local systems of teacher evaluation. His main intention is not to advocate one particular approach to evaluation, but to emphasize certain concepts, or "commonalities" as he terms them, that may become the basis for decisions. The fact that many school districts have followed his advice, and continue to do so, indicates his importance in the growth of the teacher evaluation movement.

McGreal has consistently stated that the school district must address two issues if its present teacher evaluation approach is to improve or if a new one is to be effective. First, congruence must exist between what the school district wants the evaluation system to do and to be and those things that the evaluation approach requires of the personnel involved. Particular regard must be given to the evaluation's purposes, procedures, processes, and instrumentation. All personnel involved with teacher evaluation must have relevant training to guide and develop the practices, skills, and knowledge necessary to implement and maintain the evaluation system. In the context of these sine qua non, there must be a pervasive atmosphere in the school district and schools themselves that evaluation is productive both for the individual and the organization.

Second, because evaluations necessarily lead to decisions, McGreal proposes that the procedural aspects of evaluation that lead to decisions about teachers must be clearly delineated in any evaluation design. He also stresses the importance of the quality of the relationship that exists between the supervisor and the teacher. This must be a positive relationship supported by a supervisor's skills and a teacher's commitment.

Introduction

McGreal's advice to educators about teacher evaluation has been based upon his wide knowledge of writings, research, and practices relating to models of teacher supervision, evaluation and improvement, and the organization of school districts and schools themselves. From a large amount of material, he has carefully selected those aspects that his own experiences as a teacher and his observation of practices in the field have indicated should be the most effective.

He is well aware of the pitfalls of any school district adopting wholesale a system of teacher evaluation. He therefore offers options in various broad areas of teacher evaluation characteristics, or commonalities. If the sum of these commonalities is

construed as a framework for guidance, then the school district may wish to alter its approach to teacher evaluation by choosing among the various alternatives that the commonalities offer. It shall be stated, however, that McGreal's experience has led him to believe that the perspectives he offers on some of the characteristics should be accepted. One such is his contention that different evaluation approaches are required for tenured and nontenured teachers. If an evaluation system is to succeed, it must be appropriate for the context in which it is set, so that the particular concerns, interests, and local circumstances may be closely observed and taken into consideration.

The intention of this chapter is to give a brief account of the eight commonalities that McGreal has developed for teacher evaluation. These are being gleaned partly from his writings until 1983, but principally from his book, *Successful Teacher Evaluation*, published by the Association for Supervision and Curriculum Development (ASCD) in 1983. This interesting and practical book is based on research and/or those practices that he found to be working effectively in schools.

Unless otherwise acknowledged, this chapter is extrapolated largely from *Successful Teacher Evaluation*, so that what is represented is as close a version as possible of salient points that McGreal makes in relation to his eight commonalities for teacher evaluation. As an appendix to his book, he gives an example of an evaluation system (or model) that reflects these commonalities. It is a necessary and valuable adjunct to the book. Among other things, it reemphasizes his belief that there is no need for school districts to continue teacher evaluation policies and practices that have earned the cynicism and disillusionment of both teachers and administrators.

We are indebted to Dr. McGreal for his kind permission to reproduce this appendix in full. It shows how a judicious selection from each of the eight commonalities can lead to an effective model for teacher evaluation.

The eight commonalities emphasize McGreal's belief that teacher evaluation can be both a positive and productive process.

Commonality 1: An Appropriate Attitude

Traditionally, school systems have emphasized accountability outcomes for teacher evaluation that increasingly have come into conflict with the processes designed to improve a teacher's instructional skills. The attempt to incorporate both summative (accountability) and formative (teacher development) aspects into a single evaluation approach, or model, results in a precarious position that may not suit either intention. As McGreal points out: "Trying to develop an evaluation system that walks the line between these two attitudes is extremely difficult, if not impossible" (1983, p. 2). On the other hand, a system cannot be developed that addresses

formative evaluation only, since it remains the supervisor's responsibility to ensure that satisfactory teacher competency levels are being attained. However, there seems no doubt that a school district's emphasis on a system that can separate good teachers from bad, with the ever-present threat of dismissal being an outcome of the process, has denied the potential of effectiveness of the concept of teacher evaluation.

When school districts equate teacher evaluation with accountability, directed data and documentation have to be obtained by the principal or another supervisor about a teacher's inappropriate levels of achievement. McGreal believes that school districts that have followed this approach have established a poor attitude toward evaluation. Moreover, he believes that such a use of evaluation shows a lack of basic understanding about what is needed for a process leading to teacher dismissal. Certainly, evaluation systems based on accountability promote negative feelings, which lead to an unwillingness by both teacher and supervisor to participate and to the likelihood that a teacher's competency level will not improve. By contrast, approaches that center around the concept of improving instruction are always accompanied by an acceptable level of accountability information. In this situation the prime purpose is teacher improvement rather than the meeting of organizational ends.

When it is considered that the great majority of all tenured teachers will be affected only indirectly by the outcomes of teacher evaluation during their careers, it is counterproductive for a school district to put emphasis upon a summative process. When it is further considered that legal mandates and procedures and the strength of teacher unions make the actual application of sanctions a rare event, there is further substantiation for a positive intent for evaluation.

Our chapter on the history of evaluation has shown how teacher evaluation has been negatively affected by such practices as high supervisor/low teacher involvement; criteria based on what is construed to be the health of the organization rather than the teacher's welfare; and the organization's, rather than the teacher's, criteria for judgments. The irony is that these procedures are unlikely to give any more valid basis for the dismissal of a teacher than an array of informal measures not normally associated with the term evaluation.

The right attitude is basic to the success of evaluation. Therefore, an approach should be built around attitudes directed toward improving instructional skills and procedures that complement that intention. Since evaluation must be viewed as a realistic process and since it will continue to rest on the judgments of administrators involved in the implementation of the process, it can be logically assumed that accountability measures will always accompany instructional improvement.

Once teachers and supervisors have developed a realistic understanding and attitude toward the fundamental reasons for the design and implementation of an evaluation system, acceptance follows. Teacher evaluation then has a strong chance of being successful and effective.

Commonality 2: Complementary Procedures, Purposes, and Instrumentation

McGreal sees the evaluation system as a set of required or recommended policies, procedures, processes, and instrumentation that directs the attitudes and actions of those involved in teacher evaluation. It is remarkable that some school districts that claim teacher development as the primary purpose of evaluation have in fact used methods that are counterproductive. To the extent that procedures and instrumentation fail to fall into line with policy statements, positive attitudes toward the process as a whole diminish.

As the basic test of the effectiveness of a teacher evaluation system is the relationship that exists between teacher and supervisor, McGreal suggests that an obvious starting point in developing or redesigning an evaluation system is the teacher contract. It is assumed that this will contain organizational expectations as well as statements about required teacher competency levels. A further requirement for success is that there is sensible flexibility about such aspects of the instructional process as subject knowledge in schools so that the teacher/supervisor relationship may develop unhampered by rigid prescriptions.

One reason for the apparent contradiction between policy and practice in teacher evaluation is that school districts have lacked an understanding of the range of options that are available. A knowledge of models that exist and the extent to which these can be adapted or integrated may often be the means for developing an effective system for a particular school district. McGreal cites five models and examines their usefulness.

Common Law Models. The majority of the schools in the United States have some form of common law model for teacher evaluation. Its development is usually obscure and authorship unacknowledged. Nonetheless, the characteristics of common law systems are remarkably similar, particularly since they have led to consistently negative images of teacher evaluation. Despite this, some segments of the model may fit the needs of a particular school district. The common law models are characterized by a high supervisor and low teacher involvement, by evaluation being seen as synonymous with observation, by similar procedures for tenured and nontenured teachers, by standardized criteria for judgment, and by the nature of the stated instrumentation. Such models enforce comparative judgments to be made between and among teachers. The common law models are summative evaluation.

The advantage of the common law models is that they can be used in situations where a supervisor has many teachers to evaluate, when there is little time available for training of supervisors as evaluators, and when a school district wishes to make it obvious and visible that it is striving to meet accountability demands.

Each advantage, viewed from the teacher's perspective, could be construed as a disadvantage. The common law model reinforces the traditional concept that evaluation exists for administrative purposes. A further disadvantage, in line with low teacher involvement, is that there is minimal contact time arranged between supervisors and teachers. Moreover, there is a heavy emphasis on standardized criteria, which presumes that with the identification of a finite number of criteria, all teachers shall be compared against these criteria. Related to this criticism is the fact that most criteria relevant to common law models tend to be administrative rather than teaching. Hence, common law models force comparative, rather than absolute, judgments about the effectiveness of teachers. This too easily can result in the attainment of minimal competencies but not the professional development of the individual teacher.

Goal-Setting Models. The major characteristic of goal-setting models is their emphasis on the individualized approach to evaluation. It is logical that the clearer a teacher is of what is to be accomplished, the greater the chance of success. Proponents of goal setting view it as much as a philosophy as a technique.

Goal-setting models have certain basic assumptions. If emphasis is placed upon culling out the poor teachers, such an orientation tends to equate not doing something wrong with successful teaching, whereas the focus should be on a teacher's continual growth. Priorities must be made so that the most important aspects of a particular teacher's instructional responsibilities are placed in focus. Time constraints make this imperative.

Supervision should be seen as an active process in which teachers are helped to achieve goals and grow in competency. Since teachers often perceive sections of their priority responsibilities differently from those of the supervisor or organization, goal setting must be clarified for the benefit both of the individual and organization. When priorities approximate, the result is positive and productive. In this process communication is very important.

The goal-setting model is not perfect. For instance, it cannot rank teachers, it emphasizes the attainment of measurable objectives, it too often is time-consuming, and decisions may be made on the basis of a supervisor's imperfect knowledge of subject areas. On the other hand, while focusing on correcting weaknesses and enhancing strengths, the goal-setting model promotes professional growth. It looks to the needs of individual teachers and at the same time fosters a positive working relationship between the teacher and evaluator. Moreover, it explicates expectations and sets different criteria for their evaluation. The integration of individual performance objectives with the goals and objectives of the school organization is a positive step forward.

Product Models. The product model for evaluating teacher performance has created more controversy than any other teacher evaluation approaches. It is based squarely on the use of student performance measures as the method for assessing the competency of teachers. Although there is now a diminution of the practice, by the early 1980s many states and school district authorities had established minimum competency measures and assessment programs that required or implied these measures to gauge the effectiveness of schools or teachers. Whatever arguments are in favor of using student performance data to evaluate teachers in schools, the problems in doing so are very significant.

Usually the instruments for assessing student growth are norm-referenced tests and criterion-referenced tests. Some arguments that have been advanced for use of these tests to evaluate teachers are that the student performance models are "objective," whereas those based upon such methods as "observation" are subjective. In other words, perceptible changes in student behavior brought about by a teacher's effectiveness, or lack of it, in a classroom situation is equated to true education.

While there may be surface logic to the value of using student achievement to test teacher competency, there is prevailing opinion that the inadequacy of the tests themselves, the complexity of a classroom situation, and the lack of reliable statistical measures should prevent product models from being adopted as the sole method for evaluating teachers. An open, professional discussion between administrators and teachers may lead to the inclusion of student performance as one aspect of the process.

The Clinical Supervision Model. Because clinical supervision has been extraordinarily visible and effective, many school districts have adopted clinical supervision as part of the evaluation model, or at least a major component of it. If one is seeking a positive approach to evaluation, then the clinical supervision model is a logical choice since its dominant purpose is to improve instruction. Nonetheless, as McGreal points out, there are significant definitional issues to be addressed before clinical supervision can be adopted appropriately.

Goldhammer (1969, p. 54) offers the following definition:

Given close observation, detailed observational data, face-to-face interaction between the supervisor and the teacher, and the intensity of focus that binds the two together in an intimate professional relationship, the meaning of 'clinical' is pretty well filled out.

The importance of a close and intense relationship between the teacher and the supervisor is paramount. It is assumed that teachers are professional people who require help and ways of improvement offered in a collegial rather than an authoritarian manner. Thus, acting as equals, a peer or supervisor analyzes another

teacher's performance for improvement by positive comment rather than determining correctness by admonition.

Since other chapters of this book refer to clinical supervision, further details, such as its procedures, will not be dealt with here. In particular, Madeline Hunter's approach to clinical supervision is given in detail.

In his discussion of Commonality 6 in this chapter, McGreal states that while some of the techniques inherent in clinical supervision are very useful as part of an effective teacher evaluation system, it is not appropriate to consider clinical supervision as an evaluation model.

Artistic or Naturalistic Models. Although the artistic model does not exist in local school districts, it does include some perspectives that are unique and have potential utility. The best known exponent of teaching as art is Elliot Eisner (1979), who writes that teachers, like artists, make decisions based on qualities of learning that unfold during the course of teaching. Such an artistic approach to evaluation allows both expressive and unanticipated outcomes that are of benefit to students to be analyzed.

The strength of artistic supervision may also be its weaknesses as a model for evaluation. While most teachers and supervisors see the value in discerning what is significant and subtle in student learning and the value of being able to interpret the meaning of events in the learning process, the translation of these qualities into evaluation procedures that are seen as equitable for all staff is fraught with difficulties.

Thus, the criticism of artistic or naturalistic models is centered around the lack of precision that accompanies activities relying on intuition and the analysis of the subtle qualities of learning. The time factor alone needed to train teachers and supervisors renders these models impractical, despite their inherent worth.

Summary The final test of an evaluation system is whether a relationship of mutual trust exists between the supervisor and the teacher. For this reason, procedures, purposes, and instrumentation that follow policy must all be complementary. In addition, there must be sufficient flexibility to ensure that teacher development and not teacher compliance is the prevailing attitude. There must be a clear understanding by all involved with the evaluation process that whatever choices are made among options available for consideration, the final choice has components that complement each other.

Commonality 3: Separation of Administrative and Supervisory Behavior

To establish an effective evaluation system, McGreal maintains that it is important to separate administrative from supervisory behavior. Emphasis has already been given to the fact that a successful evaluation system must have a teacher improvement orientation and a set of procedures that reflect this view. Moreover, teacher evaluation must always be seen as a *realistic* activity, one that is a natural part of the education process.

Since the vast majority of instructional supervision is conducted by administrators because their district's teacher evaluation policy requires them to do so, a minimum number of classroom visits is usually specified together with feedback, based on the district's evaluation forms and instruments. Even if the administrator wishes to evaluate for teacher development purposes, it is impossible to conduct a formative evaluation by gathering data that can only be used in making summative ratings.

This situation inhibits teacher growth, mutual supervisor/teacher trust, and the essential flexibility stressed in Commonality 2. Obviously, if the outcomes of teacher evaluation are to be positive, then procedures and instruments must be established that allow the teacher and supervisor to escape the worst effects of a poor, administratively oriented framework. While it is not expected that administrators can escape the responsibility to ascertain whether a teacher is performing according to the district's requirements, it is possible, with common sense guidelines, for the administrator to act more as an instructional supervisor than as the regulator of a district's objectives. Teachers have traditionally accepted the need for evaluation, and there is no reason why bureaucratic aspects should not be included in a teacher's evaluation for improvement, provided that the right climate and procedures are always adopted.

It is acceptable and supportable that school districts have minimal teacher performance standards encompassing such aspects as a teacher's adherence to school policy, professional attitude, personal relationships with staff and students, personal appearance, and the like. It is stressed that school districts and principals do not need to be trained to monitor the performance of teachers against these kinds of standards. Assessment goes on continuously, informally, and unobtrusively. Decisions on these observations may very well be made by an administrator who interacts in the same environment as teachers about future status, such as tenure. A special set of procedures and instrumentation does not need to be established to deal with obvious discrepancies from minimal performance standards. In line with their acceptance of the concept of evaluation, teachers are willing to accept a school district's rules and procedures provided they are appropriately handled by administrators. Therefore, there should be a separation of these kinds of decisions from

the process that is known as teacher evaluation, particularly if it is to have successful outcomes.

Some school districts are dealing with this separation by instituting different parts to their evaluation procedure. One part deals with continuous monitoring of performance, usually informally based, guided by clearly recognized minimum performance standards. Flagrant violations of these minimum performance standards cause an immediate administrative action.

A second part outlines procedures for the ways in which the supervisor and teacher will work together in the classroom, focusing on matters of instructional concern. Such a focus on techniques and a range of competencies is primarily formative in nature and based on collegiality.

In summary, a teacher evaluation scheme should do what it purports to do. Obvious instances of dereliction of duty and an unwillingness or inability to meet minimum performance standards should be dealt with administratively as they occur. The mainstream of the evaluation may then be separated from this kind of bureaucratic activity and be seen to be a commitment to improving the quality and effectiveness of classroom instruction. The implementation of teacher evaluation should then be felt to be nonthreatening. Such a climate is conducive to the acceptance of change and improvement.

Commonality 4: Goal Setting, the Major Activity of Evaluation

A characteristic of effective evaluation systems has been a development of goal setting between the teacher and the supervisor. There has been a dramatic increase of goal setting as a basic supervisory activity with the growing realization, and acceptance, by school districts that existing evaluation systems built around standardized criteria offer little or no opportunity to individualize evaluation practices.

As a formal procedure, the goal-setting process is a cooperative activity between supervisor and teacher that results in a mutually agreed upon focus for the teacher's classroom activities.

McGreal suggests three concepts of goal setting that have been used as the basic activity for evaluation systems. They are the Management by Objectives Approach (MBO), the Performance Objectives Approach (POA), and the Practical Goal-Setting Approach (PGSA). All three are based upon careful planning, implementation of what has been planned, and evaluation of the results. All three approaches have been implemented in school districts. The degree to which various approaches differ once they have been implemented depends upon the nature of the goals, the flexibility of teachers in setting goals and their measurement, and practical aspects of implementation. Thus, either selection or adaptation of one of these three approaches must be undertaken only after carefully assessing its worth to the school

system. Any selection or adaptation must complement the intention of evaluation espoused by a particular school district.

Management by Objectives Approach. This is an administrative process in which the activities of the school system are organized to achieve specific results by a predetermined date. Moreover, these results must contribute toward the achievement of any long-range objectives that the school system has promulgated.

Odiome (1965) gave this general description of MBO: It is a process whereby the superior and the subordinate managers of an organization jointly identify its common goals, define each individual's major area of responsibility in terms of the results expected of him, and use these guides for operating the unit and assessing the contribution of each of its members.

Under this system, the goals that have been chosen and that are to be achieved dictate almost all that occurs. For instance, they will determine general educational goals and plans, major aspects of the organizational structure, and goals of individual members.

As generally practiced, objectives are stated in writing, with supporting information explaining to each employee individual responsibilities for achieving organizational goals. Individuals must endeavor to determine how their personal goals can be met through the achievement of a school's organizational goals. The administrator must be concerned with time, the adequacy of checkpoints, and the flexibility to allow adjustments to be made when necessary. Educational goals are broken down into subgoals until they have meaning to all educators, and when plans are set in motion, modifications may occur until the organization's and the individual's goals reach an effective balance.

The nature of the goals and the teacher's flexibility in setting them are somewhat limited by the range of acceptable objectives emanating from the district's goals and the supervisor's goals.

The Performance Objectives Approach. Many evaluation programs follow this approach. Its originator and promoter is George Redfern who, having updated and redefined the Performance Objectives Approach (1980, pp. 21-23), states that the useful personnel evaluation program will

1. engender cooperative efforts between the person being appraised and the one(s) doing the evaluating
2. foster good communications between the parties
3. put premiums on identifying what needs improving, planning how to achieve the needed improvements, and determining how the results will be evaluated
4. promote professional growth and development of the person being appraised

5. stress the importance of evaluators becoming insightful and skilled in the art of evaluating
6. make a commitment to the proposition that the bottom line is greater effectiveness in the teaching/learning/supervising process

The POA is a cyclical process in which needs are identified, objectives and action plans set, action plans carried out, results assessed, and results discussed. There may be a need to identify further needs as a result of the assessment.

According to those who advocate POA, a prerequisite for any sound evaluation must be a clear and comprehensive definition of the duties and responsibilities of each position. These responsibilities are used as the basis for identifying needs and for beginning the cyclical process that goes on continuously. Evaluation is focused primarily on the extent to which performance objectives have been achieved.

It is not unusual in the POA approach for both teacher and evaluator, who have cooperated to identify needs, to participate in determining the success of the performance objectives. On the other hand, there is likely to be a summative aspect to the evaluation process, in which case the assessments are made by the supervisor without the involvement of the teacher. To be of any use, the supervisor will need to explain unsatisfactory ratings to the teacher. As a result of discussion between the supervisor and teacher, long- and short-range goals and objectives may be set, good work recognized, and responsibilities of both parties clarified.

One disadvantage of POA is that it may lack flexible application and local settings. For instance, the insistence on the established performance objectives having to come from a list of responsibility criteria inhibits the flexibility needed to address unique aspects of a teacher's role. This may adversely influence the supervisor/teacher relationship. Moreover, the summative rating form (as recommended by Redfern) seems an unnecessary requirement for some supervisors and one likely to harm the goodwill of teachers.

Practical Goal-Setting Approach. Like the other two approaches, PGSA is a determined attempt to focus teacher/supervision activities through a goal-setting process. PGSA, however, is more practical and less structured than the former two as it endeavors to give a realistic view of what teacher evaluation may achieve. Questions are asked, such as: What are appropriate goals? Which goals are most important? What kinds of goals are most worthwhile?

McGreal considers that there are four categories of goals that teachers and supervisors should address in normal goal-setting situations. Although these categories are listed from lowest to highest priority, the situation could arise where a goal or goals from even the lowest category must be given higher priority and addressed according to circumstantial needs.

1. Organizational or Administrative Goals
2. Program Goals
3. Learner Goals
4. Teacher Goals

Any number of specific examples, according to the context and needs, may be set under each of the four types of goals. For example, under Program Goals, the supervisor and teacher may determine that the teacher "should introduce the new reading series to the fast group in second grade"; and under Teacher Goals, it may be mutually agreed that the teacher should "work on techniques for increasing the amount and quality of student/teacher interaction."

This process of goal setting offers an excellent opportunity for personal involvement by the teacher, since the process focuses specifically on the teacher's behavior rather than on curriculum matters or learning outcomes. Goals developed from a common sense way of viewing teaching give an opportunity for supervisors and teachers to spend time together on goals that have significant bearing on student learning.

Some General Aspects of Goal Setting. Although specific goals are set, and addressed, there is no reason why broader teaching goals should not be built around these or emanate from them. As any teacher is yearly confronted with different students, texts, sets of objectives, ability levels and so on, a teaching goal built around a noncontent specific teaching skill remains with the teacher and helps address different circumstances.

It is also appropriate that supervisors and teachers accept the notion that not all goal setting must necessarily be remedial in nature. Such a flexible view opens the possibilities considerably for teacher development through evaluation.

Professional judgment, or subjectively viewing a teacher's performance, should be recognized as an acceptable and valid measurement. If an appropriate attitude toward evaluation exists and developmental plans have been sound, then the term "measurement" should be construed to mean that the supervisor and teacher together will work out methods for collecting data about each goal, so that together they make informed judgments about progress toward goals.

In the goal-setting process, it is the supervisor's responsibility to establish and maintain an atmosphere during the goal-setting conference that will allow the teacher to be an equal participant. Particularly in the PGSA approach, the supervisor must make clear how evaluation will be implemented and the kinds of evaluation instruments that will be used to gauge the degree to which goals have been met.

Negotiating goals is important. If instructional improvement is the primary purpose of evaluation, it is vital that the goal-setting activity be a mutually developed cooperative venture between teacher and supervisor. McGreal rightly

points out that this is particularly so when experienced and/or tenured teachers are involved. If conferences are to proceed constructively, it is necessary for supervisors to establish in advance the strategy they will use to make the conferences as productive as possible. Ahead of time the supervisor must work out ways to reduce threat, to make the teacher improve instructional performance, and to make it clear that both the supervisor and teacher must show a commitment to the process. Accepting this philosophy, the supervisor therefore must be willing to negotiate and compromise on issues during the goal-setting conference itself.

As in other major components of evaluation, those involved should first be given adequate training before any of the three models outlined in this commonality are adopted by school districts or modified to suit their context.

Commonality 5: Narrowed Focus on Teaching

Any teacher evaluation system developed by a school develops must center squarely on teaching itself if it is to be effective. One obvious difficulty that arises when the terms "teaching" and "evaluating" are juxtapositioned is a definitional one revolving around individual styles of teaching with their distinct characteristics.

Where school districts have evaluation systems that may be viewed as successful, there has been a decision to adopt some form of narrowed focus on teaching. Because particular perspectives on teaching are hard to find and are seldom presented in a tidy format, a degree of flexibility is built in to allow for individual differences and individual styles. Thus, the evaluation/supervision systems are based on that kind of an approach to teaching, which serves as a framework for the instructional interaction between supervisors and teachers.

There are many ways of looking at teaching that could serve as the basis for a narrowed focus. The important thing is that the school district is able to convince all staff that the selected teaching focus is appropriate and complements the evaluation policy. McGreal suggests that to achieve this, the adopted focus needs, at the least, to meet these criteria:

1. A strong empirical base
2. A close approximation to standard practice
3. A "common sense" orientation for perspectives in schools that are potentially generalized across subject areas and grade levels

McGreal goes on to suggest that the focus on teaching that seems to best meet these criteria, in terms of current teaching research and practice in school districts, is based on a combination of effectiveness research and portions of Madeline

Hunter's work. The introduction of a narrowed focus on teaching and the continued education of staff in instructional skills are essential elements in developing a successful evaluation/supervision system.

Effectiveness Research. Effectiveness research is direct application of current teaching research to improve practice. During the past decade a number of successful training and inservice packages have been generated from this source.

Effectiveness research has been chosen as the major focus by many schools because it seems to have a strong and growing research base and because research findings have paralleled accepted practice. Schools have also found attractive the fact that recommendations growing from the research are founded in common sense. Studies undertaken in a variety of settings have been reliable; that is, there has been a consistency of findings related to certain kinds of learning that pertained across subject areas in grade levels. While research in this important area must continue, there is sufficient evidence to support the view of an emergent and critical basic set of teaching skills.

Any such development does not suggest that there is any one best way of going about teaching. A good teacher will always exhibit a variety of styles according to the subject and student needs. The teacher effectiveness research strives to identify basic teaching techniques that provide fundamental skills that are applied to all levels and types of teaching.

Three important aspects of the data gathered by effectiveness research and their analysis to reach conclusions are climate, planning, and management behaviors.

Studies have begun to produce a series of results that offer a reasonably tangible definition of climate. Definitions include high levels of involvement on the part of the students, a need for teachers to plan for climate with as much diligence as they plan for the presentation of subject matter, and extended teacher/pupil contact. The skills involved in directed questioning also play a major part in climate together with the handling of incorrect responses.

Time, as a variable in learning, is another significant outcome of effectiveness research. Time must be related to planning. The way time is used by teachers and students and a relationship between directed and undirected student time have been analyzed and have given rise to changed views about lesson planning. How time is used, and should be used, by a teacher in a lesson is valuable knowledge both for the supervisor and for evaluation purposes. All involved in evaluation need to address at least the salient features of this area of teacher planning.

Significant empirical studies in the teaching effectiveness area have been directed toward the organization and management of classrooms. It has been found that where teachers have received specific training in management skills, students have achieved stronger academic attainments than in comparable classrooms with

teachers who are untrained in classroom management. Again, there is a relevance for the evaluation process in respect to a narrowed focus on teaching.

Hunter's Steps In Lesson Design. As mentioned, McGreal considers that the most successful implementations of a narrowed focus on teaching have used some of the work of Madeline Hunter and her colleagues. The very nature of the steps required in the Hunter approach demands a focus to be given. As the approach is both supervisory and clinical, the teacher and evaluator are able to anticipate a series of logical steps designed to meet particular aims.

The planning involved in the Hunter approach is illustrative of a set of practical teaching skills that form a solid and basic framework for training teachers and supervisors to focus on teaching. As Chapter 4.1 has dealt with the Hunter approach in some detail, it will not be pursued here.

Summary. Two very useful and frequently mentioned ways of narrowing the focus of teaching are provided by teacher effectiveness research and by the planning methodology developed by Madeline Hunter. Teachers and supervisors find both useful because they focus on teaching behaviors and because they adopt a common sense approach.

A narrowed focus on teaching is a vital component of any teacher evaluation system. More than any other way, it helps develop trust and credibility in the teacher evaluation process and places it in a prime position in respect to staff professional development.

A practical and realistic evaluation system that is focused on teaching and supported by teacher encouragement and training has the potential to meet the needs of both educators and of the school district as an organization.

Commonality 6: Improved Classroom Observation Skills

Classroom observation and its concomitant professional judgment form the most practical procedure for collecting formal information about teacher performance. The quality of the observations would depend very much upon the ways in which supervisors collect information and share this with teachers. Once supervisors have the motivation to wish evaluations to succeed, they should willingly undertake the necessary inservice training to improve observational skills.

Following the advice of Commonality 5 (Narrowed Focus on Teaching), supervisors should make their observations both more reliable and adequate by escaping the "wide-angled lens approach to viewing classrooms." It logically follows that the supervisor must collect descriptive data on a predetermined aspect of the teacher's performance during an observation.

Having reviewed the literature, McGreal concluded that there are four practical ways for supervisors to improve their observational skills and to use the information they collect:

1. The reliability and usefulness of classroom observation is directly related to the amount and type of information supervisors have "prior" to the observation.
2. The narrower the focus supervisors use in observing classrooms, the more likely they will be able to accurately describe the events related to that focus.
3. The impact of observational data on supervisor-teacher relationships and the teacher's willingness to fully participate in an instructional improvement activity are directly related to the way the data were recorded during observation.
4. The impact of observational data on supervisor-teacher relationships and the teacher's willingness to fully participate in an instructional improvement activity are directly related to the way feedback is presented to the teacher (1983, p. 97).

It is clear that these training guidelines are imbedded directly or indirectly in the concepts of clinical supervision referred to in the previous commonalities. As clinical supervision is often used as the basis for evaluation systems by school districts, the four rules outlined above are worthy of close attention.

There is no doubt that much classroom observation carried out in the name of evaluation has been poorly undertaken. Almost 80 percent of classroom supervision is conducted by line administrators who carry out the function mainly because it is mandated by the school district. Too often these activities are characterized by infrequent visits, attempts to view generally what is occurring in the classroom, and built-in supervisor bias or predilections.

Four Tenets of Classroom Observation. Four important tenets of classroom observation are discussed for the consideration of school districts and supervisors who wish to improve their present system of teacher evaluation or to develop a new one.

1. The reliability and usefulness of classroom observation is directly related to the amount and kind of information the supervisor obtains beforehand.

One useful sequence of events the supervisor could follow is to identify a teacher's concerns about instructions and translate these into observable behaviors. It then follows that procedures need to be worked out for improving the teacher's instruction. It is wise to give the onus of responsibility to the teacher for setting

personal self-improvement goals. When this is completed, arrangements are made for classroom observation. An observation instrument and behaviors to be recorded are selected, usually with the concurrence of the teacher being evaluated. Finally, the instructional context in which information will be recorded is clarified.

Further dimensions of this iteration would be the establishment of a contract or agreement between the supervisor and the teacher, establishing further ground rules for the observation including aspects like time, length, place of observation, and the location of the supervisor during the course of observation.

Some administrators believe that the first meeting of supervisor and teacher should be more in the way of an informational conference than a preplanning or goal-setting conference. Much will depend upon the kind of relationship that exists between the teacher and supervisor and the school climate itself, particularly in respect to the importance of teacher evaluation.

During the goal-setting conference the supervisor and teacher arrive at a mutually-agreed upon focus. A plan is then developed for working together to achieve the goals. The goal-setting conference, the observation that follows, and the postconference are essential elements of the evaluation cycle.

The accuracy of the classroom observation is directly related to the supervisor's use of a narrow focus of observation. Hyman (1975) made the following statement about observation:

Observing is much more than seeing. Observing involves the intentional and methodical viewing of the teacher and students. Observing involves planned, careful, focused, and active attention by the observer.

2. The accuracy of the classroom observation is directly related to the supervisor's use of a narrow focus of observation.

As the typical classroom is a very complex arena, it is *critical* to successful observation for the observer to be selective. It is important that, if the evaluation is to succeed, decisions about focusing on specific aspects of instruction should be made jointly by supervisor and teacher. In this respect, observation is part of the goal-setting system; the act of setting goals is a deliberate, focusing, and collaborative event.

Supervisors must learn to avoid extraneous critiques arising from the selected areas for observation, unless, of course, exceptional teacher behaviors occur that are seen to be damaging to students. Thus, the goals form the limitations for focusing the attention of the supervisor and teacher. Throwing in other feedback items will erode the very elements that make the goal setting so useful and effective.

As shown in Commonality 5, teacher effectiveness research should serve as the framework for viewing teaching. Logically, observation should be directed at

climate, planning, and management. Under climate, observers would focus on major behaviors of involvement and success; under planning, the observer could note the time factor as applied to the design sequence of a lesson; and under management, the principles outlined in Commonality 5 could appropriately be used as an observation guide.

Well-planned and executed observation can obviate the potential waste of resources arising from supervisory activities that fail to make any improvement. What counts is the *quality* and not the quantity of observation time.

3. The way data are recorded directly affects the supervisor-teacher relationship and the teacher's willingness to participate in instructional improvement.

Several different formats are available to act as observation instruments. There seems little doubt that the way that data are recorded can influence the success of the supervisory activity and the usefulness of the exercise for teachers.

Rating scales are one possible instrument. The main word of warning is that these need to be designed specifically for observations that have a particular focus in line with agreements reached during the goal-setting conference. Thus, the more specific and well-defined the items on the instrument are, the greater its utility.

There seems little doubt that the most often cited example of such an instrument is the category system designed to record a behavior, event, or interactional sequence each time it occurs. One such device widely used in school districts is the Flanders system of interaction analysis. It has been found that it is easy to learn and provides useful information on the quality and type of teacher-student verbal interaction.

4. The way feedback is presented to the teacher directly affects the supervisor-teacher relationship and the teacher's willingness to participate in instructional improvement.

In line with formative evaluation, feedback sessions occur at frequent intervals during the course of the evaluation. These should follow closely upon classroom observation. The final conference, which is summative in nature, is less advisory than judgmental by the supervisor. Since the summative evaluation is likely to be more threatening than the earlier formative evaluations for improvement purposes, it will be less productive as far as teacher development is concerned.

Some school districts have instituted two different kinds of evaluation: one for teacher development and the other to make summary decisions about areas like tenure or even dismissal.

In general terms, feedback should be focused on the actual performance of the teacher rather than on a personality. Emphasis must be on observations rather than assumptions or inferences and based on descriptive rather than evaluative comments. The specific rather than the general should be the basis for discussion, and this should lead to the sharing of information rather than judgmental advice. Care must be taken that feedback contains what the teacher can use and manage rather than a pouring out of all the information that the supervisor has gathered.

The final summary is useful to both teacher and supervisor. Written evaluation should not be made by the supervisor until completion of this final conference.

The Written Report. This should be made as positive as possible. Essentially, it must be based on the agreed-upon goals and the kind and quality of conversations that followed in the postobservation conference.

If value judgments are made they should be supported by example, anecdote, or description. The final report, which may be termed summative in nature, must not contain surprises for the teacher as all salient features should have been thoroughly discussed with the teacher during the final, or earlier, conferences.

Observation, the dominant method for collecting information about teachers in the classroom situation, may be very reliable. Much will depend upon the training and developed skills of the supervisor in focusing upon aspects of a teacher's instructional behavior.

Commonality 7: Use of Additional Sources of Data

Observation is only one way of collecting information about teaching. Among other alternatives are self-evaluation, peer evaluation, parent evaluation, student evaluation, student performance and an artifact collection. While each may be useful under certain circumstances, McGreal selected two methods as being more useful than the others. One of these is student evaluation as a source of information or, more accurately, student descriptive data. The other is the compilation of an artifact collection, which may include study guides, question sheets, homework assignments, experiments, tests, and the like.

There is no doubt that teachers are evaluated directly and indirectly by parents. The problem of their involvement is a political one. If parents are to visit schools and be involved in other ways in activities that may be construed as teacher evaluation, then it should be done in the positive context of a public relations structure. Any full-fronted involvement by parents in teacher evaluation could well destroy the positive and sensitive nature of the process.

If peers are involved in a teacher's evaluation in positive, mutually supportive, and nonthreatening ways, there should be useful outcomes. In the past there has

been almost unanimous objection to the concept of observation and evaluation by peers. The teacher being observed too often becomes embarrassed because weaknesses or deficiencies are seen by a colleague and are subject to critical comment. On the other hand, it would appear that peers are in a good position to provide both reliable and valid evaluation of each other. If peer supervision can occur naturally as part of the planned system—and this does occur in some school districts—it can be a valuable part of the total evaluation process.

There appears little doubt that information gathered about student learning is an important source of information about the effectiveness of a teacher. However, like so many other issues concerned with evaluation, the logic of the idea is often subsumed by practical and political considerations. References were made earlier in this chapter to research data that have indicated that the imperfection of standardized testing is both a powerful and logical argument against student performance data being used for teacher evaluation.

Most writers and practitioners of teacher evaluation would give a stronger place to self-evaluation than does McGreal. Ultimately, any teacher improvement must be self-motivated, despite the various supports offered by the school district and by the supervisor. While some teachers clearly are diffident about self-evaluation and indeed would underrate themselves, most aspire to enhance their professional status and welcome self-assessment as the most acceptable form of evaluation.

Self-evaluation statements should not remain the preserve of the teacher. These need to be shared and, as suggested in Chapter 4.7 depicting the Shinkfield model, form an essential basis for collegial decision making about goal setting. Thus, self-evaluation leads toward cooperative, professional interaction between the evaluator and the teacher.

Student Evaluation. Most research in the area of student evaluation of the teacher's performance has been at the tertiary [university and college] level. Elementary and secondary teachers have traditionally felt uncomfortable with the concept, perhaps because they lack faith in the student's ability to accurately assess their performance.

Much depends, however, on the kind of data collected and how these are used.

Under the right circumstances it is possible for student evaluation to be seen as both acceptable and useful. As McGreal points out:

The major ingredient for the successful use of student evaluations is the acceptance of the idea that students are much more reliable in describing life in the classroom than they are in making evaluative judgments of the teacher (1983, p. 134).

This view places the emphasis on classroom activities and student learning, rather than on teacher personality.

There is no suggestion that student feedback should be anything but an alternative or additional source of information to classroom observation by the supervisor. Wisely constructed student evaluation may well be a sound and complementary adjunct to observational procedures.

Artifact Collection. The value of the compilation of an artifact collection is only just beginning to be fully realized. Classroom observation, no matter how well focused, can only capture certain aspects of what a teacher has planned and indeed is undertaking in the classroom.

If observation is supported by a collection of artifacts made or gathered by the teacher and if these are reviewed, analyzed, and discussed, a far more complete and more fair perception of ways in which the teacher is helping student learning may be perceived. When it is realized that the majority of a student's day in school is spent in seat work and related activities, collecting and reviewing a teacher's artifacts throws important light on that teacher's effectiveness. Moreover, concepts of classroom planning that go beyond the traditional lesson plan can be discerned through the collection, presentation, and subsequent discussion of artifacts.

Commonality 8: A Training Program Complementary to the Evaluation System

McGreal emphasized that an evaluation system is effective only if all those who are to be involved are adequately trained. In fact, he stated that the success of evaluation system is directly inproportional to the quality of the program offered and the provision of adequate time to ensure that policies, procedures, and processes are all fully understood.

It has been traditional for school districts and schools themselves to give scant regard to training for the evaluation process. This inadequacy may well have arisen because too often lip service only has been given to teacher evaluation and its importance in the full educational development of the school district. Moreover, such school districts are unlikely to offer credence or support to the evaluation process leading toward the professional growth of staff.

The various commonalities presented in this chapter have all led one way or another to the conclusion that a training program is essential. Although considerable time may be spent with supervisors on the irresponsibilities with goal setting, observation techniques, conferencing, and feedback skills, both administrators and teachers should be given approximately the same training. One obvious benefit of this emphasis on training is that the concept of evaluation and its importance to the school district are made visible.

Any training program must be well structured so that the knowledge and skills necessary for the implementation of an evaluation system may be fully addressed. The entire staff may require a total of up to eight hours to encompass the introduction to the system, teaching focus, goal setting, data collection methods, and summary discussions. In addition, supervisors will need a full day to grasp their specific skills training.

Thus, although it does not take a significantly large amount of time to prepare participants, the quality of what is offered is of paramount importance. Focus must be brought to bear on the teaching/learning process and on the enhancement of teacher/supervisor relationships. These areas touch upon almost every facet of effective schools.

Following the training program, steps are taken to develop and implement a new evaluation system or to revise the extant one. A committee may be selected from teachers and administrators by virtue of their competence, their involvement in the teacher association, or the respect with which they are held by their colleagues. If it is thought necessary, an outside consultant should be employed to suggest alternative approaches based on experience, research, and current successful practices.

The process from establishing the committee to developing the system and to providing training need not be a lengthy one. In fact, the school district and interested community should see one step evolve from the previous set of decisions.

Conclusion

McGreal supplies an excellent appendix to his book *Successful Teacher Evaluation*, which gives an example of an evaluation system that reflects the eight commonalities that he has presented.

He emphasizes that aspects of this example are not definitive and that a particular school district must select from among the options available those aspects that suit its own context and predilections about education.

As the appendix shows, the example that McGreal offers is simple, yet sophisticated in that it encompasses the major concepts that he has demonstrated to be useful for a successful evaluation system.

The example commences with a philosophic statement and then goes on to give detailed information under these headings:

1. Minimum Performance Expectations
2. Improvement of Instruction
3. Attachment (which includes criteria for teacher effectiveness, goal setting, and techniques for determining teacher effectiveness)

There is a growing realization that there is no area in education that has more potential for the improvement of instruction and indeed the improvement of schools themselves than a successful teacher evaluation system. The 1983 Federal Commission on Excellence in Education document, *A Nation At Risk*, has given a further imperative for effective teacher evaluation. Therefore, despite declining numbers of students in age cohorts and more limited resources, it is essential that schools and school districts develop a supervision/evaluation system that uses the skills of existing staff to the extent possible so that the quality of instruction can be enhanced.

References

- Eisner, E. (1979). *The educational imagination*. New York: MacMillan.
- Goldhammer, K. (1969). *Clinical supervision*. New York: Holt, Rinehart, and Winston.
- Hyman, R. (1975). *School administrator's handbook of teacher supervision and evaluation methods*. Englewood Cliffs, NJ: Prentice-Hall.
- Iwanicki, E. (1981). Contract plans: A professional growth-oriented approach to evaluating teacher performance. In J. Millman (Ed.), *Handbook of teacher evaluation*. Beverly Hills, CA: Sage.
- McGreal, T. L. (1980, February). Helping teachers set goals. *Educational Leadership*, 37, p. 414).
- McGreal, T. L. (1982, January). Effective teacher evaluation systems. *Educational Leadership*, 39, p. 303.
- McGreal, T. L., (1983). *Successful teacher evaluation*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Odiorne, G. S. (1965). *Management by objectives*. New York: Pitman.
- Redfern, G. (1980). *Evaluating teachers and administrators: A performance objectives approach*. Boulder, CO: Westview Press.

Appendix: An Example of an Evaluation System That Reflects the Commonalities of Successful Systems¹

Philosophy

The parents, school board members, and staff of _____ are committed to the continuation of the district's strong educational program. An effective

teacher evaluation system that focuses on the improvement of instruction is an important component of this instructional program.

While this primary focus of evaluation is to improve instruction, teacher evaluation requires teachers to meet the established performance expectations. This process must be continuous and constructive, and must take place in an atmosphere of mutual trust and respect. The process is a cooperative effort on the part of the evaluator and teacher. It is designed to encourage productive dialogue between staff and supervisors and to promote professional growth and development.

1. Minimum Performance Expectations

An integral part of both tenured and nontenured staffs' employment in the school district is continuous appraisal by their supervisors of their ability to meet minimum performance expectations. As appropriate to the various jobs performed by staff members, the minimum performance expectations include, but are not necessarily limited to, the following:

1. Meets and instructs students at designated locations and times.
2. Develops and maintains a classroom environment commensurate with the teacher's style, norms of the building program, appropriate to the classroom activity, and within the limits of the resources provided by the district.
3. Prepares for assigned classes and shows written evidence of preparation and implementation on request of the immediate supervisor.
4. Encourages students to set and maintain acceptable standards of classroom behavior.
5. Provides an effective program of instruction based on the needs and capabilities of the individuals or student groups involved. This should include, but not be limited to:
 - Review of previously taught material, as needed.
 - Presentation of new material.
 - Use of a variety of teaching materials and techniques.
 - Evaluation of student progress on a regular basis.

¹ Special thanks are extended to the Penn-Harris-Madison School Corporation in Mishawaka, Indiana; Pikeland School District, Pittsfield, Illinois; Monticello Public Schools, Monticello, Illinois; West Aurora Public Schools, Aurora, Illinois; and Palos District #118, Palos Park, Illinois, for permission to reproduce parts of their teacher evaluation procedures.

6. Correlates individual instructional objectives with the philosophy, goals, and objectives stated for the district.
7. Takes all necessary and reasonable precautions to protect students, equipment, materials, and facilities.
8. Maintains records as required by law, district policy, and administrative regulations.
9. Assists in upholding and enforcing school rules, administrative regulations.
10. Makes provision for being available to students and parents for education related purposes outside the instructional day when necessary and under reasonable terms.
11. Attends and participates in faculty, department, and district meetings.
12. Cooperates with other members of the staff in planning instructional goals, objectives and methods.
13. Assists in the selection of books, equipment, and other instructional materials.
14. Works to establish and maintain open lines of communication with students, parents, and colleagues concerning both the academic and behavioral progress of all students.
15. Establishes and maintains cooperative professional relations with others.
16. Performs related duties as assigned by the administration in accordance with district policies and practices.

The appraisal of these minimum expectations will typically be made through a supervisor's daily contact and interaction with the staff member. When problems occur in these areas, the staff member will be contacted by the supervisor to remind the staff member of minimum expectations in the problem area and to provide whatever assistance might be helpful. If the problem continues or reoccurs, the supervisor, in his or her discretion, may prepare and issue to the staff member a written notice setting forth the specific deficiency with a copy to the teacher's file. In the unlikely event that serious, intentional, or flagrant violations of the minimum performance expectations occur, the supervisor, at his or her discretion, may put aside the recommended procedure and make a direct recommendation for more formal and immediate action.

II. Improvement of Instruction

This part of the appraisal program uses a positive approach to stimulate self-improvement as well as creating a continuous focus on improved instruction and/or the delivery of instructional support. The supervisor and the staff member share the

responsibility for this procedure. The fundamental supervisory activity of this program is the development of specific teaching or direct job related goals between the staff member and the supervisor. Part A in the Attachment discusses current teacher effectiveness research that should serve as the basis for most teacher goal setting. This appraisal plan is formative (data gathered for the purpose of improving job performance) and bilateral in nature. *Its purpose is to focus on the delivery system of instruction, with the staff member and supervisor working together to increase teaching effectiveness and student learning.*

Required and Recommended Procedures for Part II

1. All nontenured staff will be involved in the goal-setting process each year.
2. All tenured staff will be involved in the goal-setting process every second year. Participation the first year will be determined by alphabetical order in each building. A tenured person may participate in the goal-setting process in successive years if deemed necessary or useful by the supervisor or staff member.
3. This part of the appraisal program will be conducted by the immediate supervisor of the staff member or by a designated representative. Itinerant staff will be appraised by a designated "home" supervisor.
4. The goal-setting conference should be held as early in the year as possible, preferably by October 15 (each year for nontenured, every other year for tenured).
5. There are three basic parts to the goal-setting conference:
 - Establishing goals:
Nontenured staff. During the conference the supervisor should take the lead in establishing goals. The recommended guidelines for goal setting as described in the Attachment, Part B, should be used.
Tenured staff. Tenured staff are expected to play an active role in establishing goals. The recommended guidelines for goal setting as described in the Attachment, Part B, should be used. If agreement cannot be reached on the goal(s), the supervisor will have final responsibility.
 - Determining methods for collecting data relative to the goals: As each goal is established, the means for collecting data to determine progress should be determined by the supervisor and the staff member. The three most recommended methods for collecting data are discussed in the Attachment, Part C.
Nontenured staff. Each nontenured staff member must be involved in the use of all three of the recommended methods. Those staff members not involved in direct instruction would be excused from this requirement.

- Observation—each nontenured teacher must be observed in the classroom throughout the year.
- Artifact collection—once during the school year, all artifacts used or produced during the teaching of one unit will be collected and reviewed with the supervisor.
- Student descriptive data—once during the school year, information will be gathered from at least one class of students regarding their perceptions of life and work in the classroom.

Tenured staff. The means for collecting data regarding progress should be discussed and agreed upon by the staff member and the supervisor. The method selected should be appropriate to the goal. There are no specific requirements as to the type or frequency of methods. In those instances where agreement cannot be reached, the supervisor has the final responsibility.

- A written description of the goal-setting conference:

Part D in the Attachment provides a standard form to be used by the supervisor for writing a description of the goal-setting conference. It should be written during or immediately after the conference and shared with the teacher. It should be submitted at the end of the appraisal period as part of the final write-up.

6. During the actual appraisal period (following the goal-setting conference to the time of completion of the final appraisal report) records of the interactions, contacts, activities, and so forth between the supervisor and the staff member should be kept. These would include such things as dates and summaries of observations; records of student evaluations; findings from artifact reviews; and summaries of other training contacts with the staff member. It is generally the recording of any and all contacts or data that are appropriate to the methods agreed upon by the supervisor and the staff member during the goal-setting conference.

The Final Appraisal Conference should be held at the end of the appraisal period (the first week in March for nontenured staff, by the third week in May for tenured staff). It is the concluding activity in the appraisal process. The form provided in the Attachment, Part E, should be used to provide a summary of the conference. The highlight of the conference should be the joint discussion of the year's activities, the implications for future goal setting, and continued self-growth. The summarizing write-up should be done during the conference or immediately afterward. The summary should be a clear reflection of the discussion during the conference and be shared with the staff member for his or her signature and optional comments.

Part A: Criteria for Teacher Effectiveness

The basic criteria to be used in setting goals during the initial supervisor-teacher conference is based on current teacher effectiveness research (teacher behaviors related to student achievement). The concepts presented below represent a summary of current research (1981) and should be used as guidelines whenever possible. These statements are presented as a framework for looking at classroom practices and are not presented as a checklist of required practices. In those instances where the person being evaluated is not involved in the direct instruction of students, it is assumed that other direct job-related criteria would be more appropriate.

Classroom Climate

1. Positive motivation is evidenced.
2. A focus on student behavior rather than personality is reflected.
3. Classrooms are characterized by an environment in which all the students feel free to be a part of the class.
4. There is a high degree of appropriate academic praise for all students.
5. Concern for increasing the percentage of correct answers given by students in class and on assignments while at the same time holding expectations realistically high is apparent.
6. The teacher demonstrates active involvement and visible leadership.
7. The teacher gives the impression of enjoying working with students and reflects respect for them as individuals.

Planning

1. All pupil contact time is planned.
2. Teaching unit plans generally include the following:
 - Clearly identified long-range goals and short-term objectives.
 - Materials and methods to be used, showing a variety of ways to illustrate information.
 - Special supplementary resources when appropriate (such as library, field trips, resource people).
 - Provisions for students to have guided and/or independent practice.
 - Methods to be used in checking for student understanding, getting sufficient feedback.

3. Daily written lesson plans are detailed enough for teachers' and/or substitutes' use.
4. Objectives of instructional plans relate directly to the objectives of the District's adopted curriculum, using adopted program materials (manuals, course descriptions, student texts, recommended supplementary materials).
5. Instructional plan demonstrates an understanding of the content and an awareness of the variety of ways skills can be learned.
6. Pupils' subject matter strengths and weaknesses and academic, social, emotional and physical needs are identified and planning takes these into account.

The Teaching Act

1. Explanations, demonstrations, practice, and feedback are presented so that the students can comprehend and retain what is being taught. Includes the following steps:
 - Establishing mental set at the onset of the lesson, e.g., providing students cues that arouse interest.
 - Teacher clearly stating to the students the objectives of the lesson.
 - Teacher or students illustrating what is to be learned.
 - Checking for student understanding.
 - Providing students with guided practice.
 - Providing students with independent practice.
2. Varied groupings, methods and materials used are based on the needs of the students and objectives of the lesson.
3. Emphasis is placed on providing high percentages of academic engaged time.
4. Recognition is given to the importance of the appropriate use of a direct instruction teaching model: Keeping students on task; direct supervision skills; quality of seat work.
5. All non-direct teaching activities are monitored for their usefulness and appropriateness (i.e., seat work assignments, homework, tests, and quizzes, use of interest center, independent study, activities, individualized instruction activities).

Management Skills

1. Teacher planning maximizes student on-task time.
2. Limits of student behavior are clearly defined, communicated to students and consistently monitored.

3. Teacher monitors rest of the class while working with small groups and individuals.
4. Teacher organizes and arranges classroom so as to facilitate learning and to minimize student disruption.
5. Transitions from one area of teaching to another are made smoothly and demonstrate pre-planning.
6. All students are treated in a fair and consistent manner, taking individual needs into account.

Supplement to Criteria for Teacher Effectiveness. The following definitions and examples are intended to clarify terms and indicate the intent of concepts. Examples should not be considered the limits of the expectations. No attempt is made to provide a rationale for the criterion. The numbers and letters are keyed to the above "Criteria for Teacher Effectiveness."

A.1. "Positive motivation":

- Provides opportunities for right answers.
- Responds to wrong answers with supporting techniques, such as clarifying question,
- Chooses and phrases questions that facilitate correct answers.

A.2. "Focus on behavior":

- Encourages students to volunteer answers.
- Uses students' responses and ideas.

A.3. "Environment in which students feel free":

- Uses varied questions so that all students have a chance to be successful in their responses even though some questions may well be beyond some students.

A.4. "Appropriate academic praise":

- Plans situations so that all students have the opportunity to earn praise for academic effort and accomplishment.
- Plans assignments to promote a high degree of success yet maintain a moderate challenge.
- Emphasizes what is correct about students' work rather than only noting errors.

A.5. "Percentage of correct answers":

- In daily work and class participation, average and below average students have at least 70 percent correct answers; more able students at least 80 percent correct.
- Wrong answers are probed to success, especially with average and below average students.

A.6. "Active involvement and visible leadership":

- Responsive and involved verbally and nonverbally.
- Regardless of activity, is involved—explaining, leading, or participating in discussions, observing individuals' work, interacting with individuals or small groups.
- Is not grading papers, reading, planning for another class, talking to other than a class member in the classroom or hall.
- Recognizes and reinforces appropriate behavior and clearly sets the tone for the class.

B.1. "Contract time":

- The Period during which the teacher is responsible for the instruction of pupils.

B.2. "Teaching unit plans":

- Plans for a major topic or section of student work extending over several days or weeks; usually relatively short at lower elementary to extended at the secondary level.

B.6. "Planning takes these needs into account":

- Formal or informal pretesting used to assess pupils' competence.
- Uses supportive personnel for identification, diagnosis, planning, and identification as appropriate.

C.1. These seven elements of a lesson may not all occur in a given period, but the sequence is generally applicable when dealing with a new or extended skill or concept. Omission of a step should be conscious for an educationally sound purpose.

- "Mental set":
 - Focusing attention on the concept or skill to be studied; in this sense, more than just getting the attention of the class.
- "Presenting information":
 - Teacher or student explanation or demonstration.
 - Assigned readings.
 - Audiovisual material.
 - Resource persons.
- "Checking for student understanding":
 - Questions asked of a sampling of the class.
 - Sample exercises on the chalkboard or overhead projector are done by students.
 - Typically, "Any questions?" or "Do you understand?" are *not* sufficient.
- "Guided practice":
 - A few examples are done independently by students in class with the teacher checking each to ensure individual understanding; explaining and clarifying when necessary before assigning independent practice.
- "Independent practice":
 - Application of skills or concepts by individuals after teacher has ensured their understanding through guided practice; may be long- or short-term, in school or homework.

C.3. "Academic engaged time":

- Time when pupil is actively involved in academically appropriate activity; listening may or may not be academic engaged time.

C.4. "Student on-task time":

- Time when the student is directly involved in academic work related to the lesson or other specified objective; similar to academic engaged time, but could include nonacademic activities; student works on what he or she should be working on.

Part B: Goal Setting

Both the supervisor and the staff member have a responsibility to make the goal setting conference as productive as possible. The supervisor, while maintaining ultimate responsibility for the final product, must actively involve the staff member in the conference. In most instances, the final goals should be the outgrowth of a cooperative activity. (In working with nontenured staff, the supervisor will normally assume a more directive role in goal setting. With tenured staff, the supervisor's major functions would tend to be as a clarifier and facilitator. When agreement cannot be reached, the supervisor maintains final responsibility.) The staff member is responsible for coming to the conference prepared to openly and positively discuss areas that are of particular concern or interest. Both parties share the responsibility of approaching the conference and the entire activity with a positive attitude and a willingness to participate fully.

Number of Goals. The number of goals established between the staff member and the supervisor is less important than the form and substance of the goals. In most cases, the number would range between one and four, with the number being determined by the relevancy and the time and energy required.

Goal Priorities. Under normal conditions, it is recommended that goals be established in accordance with their potential impact on student learning. The following priorities should be used as guidelines in determining the appropriateness of goals. However, there are instances when any one of the four types may be relevant and necessary depending on unique conditions.

1. *Teaching Goals*—goals built around teacher behaviors or worker behaviors that are directly related to student outcomes. The outline of the teacher effectiveness research in the Appendix-Part A should serve as the basis for setting teaching goals for the regular classroom teachers. Other instructional support personnel should consider direct job-related activities as falling under this heading.
2. *Learner Goals*—goals that relate directly to solving a specific learning activity or improving some particular student deficit.
3. *Program Goals*—goals that relate to curriculum areas, course outlines, articulation activities, materials selection, etc. It is assumed here that there are numerous ways for staff to get involved in programmatic efforts other than using the supervision system.
4. *Organizational or Administrative Goals*—goals that deal with specific administrative criteria such as listed in the minimum standards description. It

is assumed that only in the case of continuing problems in this area would the goal-setting procedure be used to help improve the situation.

Measurability of Goals. Part C in the Appendix lists the preferred options for measuring progress towards meeting the goal(s). The key to this activity during the conference is a cooperative effort between the supervisor and the staff member in arriving at a method that fits each goal. Certain goals may be so unique that they force the supervisor and staff person to creatively design a method for assessing progress. This is perfectly acceptable. It is to be remembered that subjective judgments made by the supervisor and the staff person after the method(s) have been applied are clearly acceptable forms of measurement. This allows us not to have to confine our goals to only those things that are measurable by traditional, empirical standards.

Part C: Techniques for Determining Teacher Effectiveness

Several techniques can be employed to formatively collect data about classroom instruction.

Formal Observation. Observing the teacher in the classroom is a basic and important way of determining teacher effectiveness. Formal observation will be made throughout the school year with either the teacher or supervisor initiating the formal observation process. To increase the reliability of the information gained through the formal observation, the following procedures will be required of all formal observations.

1. A pre-observation conference is required for each formal classroom observation to help the teacher and supervisor determine the primary focus of the observation. In the pre-observation conference the following information is to be discussed.
 - Specific area of Teacher Effectiveness Criteria that will receive primary emphasis during the observation.
 - Student outcomes to be achieved by the lessons.
 - Methods teachers will use to help the students achieve the lesson objective.
 - Behavior students will display that will indicate their successful achievement of the lesson objective.

2. The pre-observation conference may be held at any time prior to the observation. The formal observation form is to be used to record information collected during the formal observation process.
3. A description of the observation will be given to the teacher within a reasonable time prior to the post-conference.
4. A post-observation conference will be held following each classroom observation with such conferences being conducted within a reasonable time following the observation—usually not more than two school days. Information determined in the observation and pre-observation conference will form the basis of discussion in the post-conference.

Artifact Collection. An important appraisal alternative to the formal observation process is artifact collection. Artifacts would include such things as lesson plans, unit planning materials, tests, quizzes, study guides, worksheets, homework assignments and other materials that affect or relate to instruction. The Teacher Effectiveness Criteria will serve as a basis for determining the quality and appropriateness of classroom artifacts. A conference may be scheduled for the purpose of mutually appraising instructional artifacts with requested data being presented to the supervisor at least one day prior to an arranged conference. All artifacts reviewed in the conference will be returned except those that have been mutually determined to be used for the preparation of the final appraisal report.

Student Evaluation. Great insight can be gained related to instructional effectiveness and effective classroom procedures by asking students for their reactions and perceptions to questions aimed at producing descriptive information about the classroom and the instruction in that classroom. The purpose of any such appraisal is to obtain descriptive data about instruction and not to rate the teacher. Such information will be mutually reviewed by the teacher and the supervisor to determine the level of instructional effectiveness in the classroom. Any written information, forms or notes used or made in employing this technique as a data source shall be shared solely between the teacher and the supervisor. The results of this appraisal technique would not be included as part of the teacher's Annual Appraisal Report unless both the teacher and the supervisor mutually agree to do so. (Various student evaluation instruments will be made available through the Office of the Assistant Superintendent.)

Part D.

Staff Member _____ Supervisor _____

School _____ Date _____

PRE-APPRAISAL CONFERENCE

- A. Establishment and Monitoring of Performance Goals (attach additional material as needed).

**Performance Goals for
Appraisal Period**

**Means for Measuring the Degree
to Which the Goal was Reached**

- B. Additional Comments Relevant to the Conference

*Part E.***FINAL APPRAISAL REPORT**

Staff Member _____ Supervisor _____

School _____ Date _____

A. *A Summary of the Appraisal Process*B. *General Follow-up Recommendations*C. *Remarks by the Staff Member (optional)*_____
Signatures indicate completion of the process, but not necessarily agreement.

Teacher _____ Date _____

Supervisor _____ Date _____

Edward Iwanicki: Contract Plans—A Professional Growth-Oriented Approach to Evaluating Teacher Performance

Over the years, the concept of contract plans for teacher evaluation has been widely criticized. In fact, the weaknesses of the approach have received as much emphasis in the literature as its strengths. One who currently writes in favor of the approach is Edward Iwanicki of the University of Connecticut School of Education. Although he acknowledges the potential flaws in contract plans, he nonetheless shows that the approach is both feasible and defensible, provided that all concerned personnel have a professional interest in its success and that, concomitantly, the planning, organization, and implementation of a selected approach are thorough.

Of those who have written about contract plans for the professional growth of teachers, Iwanicki is possibly the most evenhanded and practically oriented. This chapter consists primarily of a version of the chapter written by Iwanicki in *Handbook of Teacher Evaluation* (edited by Jason Millman, 1981). Similarities with the models presented in this book by Hunter (Chapter 4.1) and Shinkfield (Chapter 4.7) in respect to aspects of clinical supervision are interesting, since all three approaches have been developed independently; but all three stress improved learning by students as the motivation for improved teaching.

Introduction

Critics of teacher evaluation models and approaches have gone to some lengths to express the opinion that contract plans to improve teacher performance have a strong chance of success only if they are devised to meet the professional needs of teachers. If teachers are to be ranked for promotion or to have tenure confirmed, then other techniques should be investigated.

The essence of contract plans as a process is cooperation between teacher and evaluator as the following evaluation cycle unfolds:

1. Teacher performance is reviewed.
2. Priority areas for improvement are identified.
3. An improvement plan containing performance objectives is developed for each priority area.
4. The improvement plan is implemented and monitored.
5. The impact of the improvement plan on teacher performance is evaluated.

For contract plans to succeed it is assumed that teachers are professional people who seek to improve their performance because they realize that societal changes will inevitably change student, and therefore curriculum, perspectives.

Performance objectives form part, although not necessarily all, of the basis for the contract plan approach. Teachers may also be evaluated on how they perform these vis-à-vis their stated job description.

Two Approaches: Management by Objectives and Clinical Supervision.

Iwanicki points out that contract plans may be implemented on the basis of the assumptions and procedures underlying management by objectives (MBO) or those of clinical supervision or, indeed, both.

The management by objectives approach compares performance against priority objectives set by the organization in line with its mission, purpose, and long-range aims. When a clinical supervision approach is adopted, a staff member's performance is analyzed against his or her particular role in the school organization. In this situation, performance objectives play a two-fold part: they must strengthen the teacher's performance in the role that he or she has to play, and they must meet the professional needs of that teacher.

It is Iwanicki's opinion that a preponderance of the MBO approach, which tends to give priority to the organization, or a preponderance of the clinical supervision approach, which will strengthen professional growth to the possible disadvantage of a grasp of the institution's total educational program, are not recommended. The MBO approach tends to give teachers the feeling that they have been coerced into developing objectives defined by the administration. Sergiovanni (1974), however, has stated that in highly synergistic organizations it would be possible to accommodate either largely MBO-oriented or clinical supervision-oriented approaches to contract plans with positive results.

Iwanicki quite rightly points out that the central issue is not which is the better of the two approaches provided that the outcome is improvement in the quality of educational programs. The adopted approach, or combination approaches, must contribute to the needs of the organization as well as the personal needs of individuals. The success of schools is judged from the quality and extent of student learning. Teachers who are given the means to develop professionally in a healthy organization are the main contributors to a school's success.

Iwanicki's Prerequisites for Effective Implementation of Contract Plans

Any personnel evaluation process in any organization will succeed only if it has a firm foundation. Planning and organization are essential to secure teacher coopera-

tion and commitment and acknowledgement of their legitimate vested interests. Iwanicki stresses this point: "Unless staff accept the fact that effective teacher evaluation is an integral part of their professional responsibility, it is most difficult to build a commitment to Contract Plans" (1981, p. 203).

It is all too early to look skeptically, or lazily, at the need for thorough planning. The enticement to adopt an evaluation program that apparently has worked successfully elsewhere is very strong. Apart from the fact that other evaluation programs may be inappropriate for a whole host of reasons, the value of local involvement in the development of an evaluation scheme must never be underestimated.

School districts and individual schools contemplating personnel evaluation should commence their plans for teacher evaluation with the realization that their own staff, curriculum, and material sources are largely unique and that some extent of invention will be essential for the implementation of a successful teacher evaluation program. While the modification of a known (and successful) program should not be discounted and very well might be beneficial, all concerned with a local evaluation must be assured that the adopted approach is consistent with the conclusions they have reached about the philosophy and intentions underlying their own plan for teacher evaluation. An individual school's philosophies and goals must provide the foundation for a teacher evaluation program if it is to be effective.

It should not be assumed, however, that a school will maintain its educational hopes and intentions unchanged over a period of time. A review of all major aspects of a school should be made from time to time, and in conjunction with such a review, a restatement of the philosophies and goals that form the criteria for judgments and decisions related to teacher evaluation. A school's goals and philosophies must be stated, or interpreted, in measurable or observable terms through the development of program objectives. A representative statement of these objectives conveys the outcomes that the organization hopes to achieve through the professional diligence of its personnel.

Developing Job Descriptions. Once the philosophic and goal statements have been completed, job descriptions may be developed specifying the performance criteria expected of teachers responsible for realizing intended outcomes. Job descriptions are therefore **essential** to the contract plan approach. They will describe the behaviors expected of a teacher placed in a particular circumstance and given a particular responsibility. Moreover, they become the criteria used to evaluate the teacher's performance. As such they must identify closely with competency areas that the literature has shown to be crucial to the effective performance of a teacher undertaking particular professional tasks within the cultural milieu of the school. Job descriptions should be sufficiently specific to enable descriptive criteria that will be used to evaluate their performance to be

understandable to teachers. A globally stated job description will be more open to different interpretations and ambiguity than a detailed job description related to the chosen objective or behavior.

Iwanicki cites an example from the Enfield, Connecticut, Public Schools to indicate that a global criterion statement may effectively be translated into specific criteria that indicate to the teacher what the organization's expectations are in respect to the areas of colleague and parent interaction. The example indicates that it is both easier and more sensible to identify where improvements are expected by using specific criteria in conjunction with the global criterion statement.

Global Criterion Statement. The teacher interacts effectively with colleagues and parents.

Specific Criteria. The teacher

1. is willing to cooperate with coworkers by sharing ideas and methods of instruction
2. exhibits ethical behavior toward fellow teachers and coworkers
3. attends committee and faculty meetings
4. seeks assistance, advice, and guidance as necessary from colleagues and/or specialists
5. confers with parents, when necessary and possible, to foster a constructive parent-teacher relationship
6. involves parents in class-related activities when appropriate (1979, p. 44)

It is interesting to note that the majority of competencies presented in most job descriptions are pertinent to all teachers. Those with more limited application are determined by the background of students being taught, curriculum intentions, grade level, and possibly, the extent of teacher preparation.

Connected to the development of job descriptions must be the understanding of accountability relationships. It is essential that teachers know to whom they are accountable in the evaluation process, who is primarily responsible for evaluating their performance, who else may have a role in the evaluation process, and for what reasons. In a large, complex organization these roles may become difficult to define. Nonetheless, an attempt should be made to do so. It is possible that subject specialists may necessarily be included to form an evaluation panel, as there is the increasing realization that school administrators may no longer have sufficient expertise to conduct an evaluation alone. Therefore, heads of departments and curriculum resource personnel are playing a more active role in the teacher evaluation process.

Defining the Purposes of Evaluation. Bolton (1973) identified the following purposes for teacher evaluation:

1. improvement of instruction
2. rewarding superior performance
3. modification of assignment
4. protection of individuals and the organization
5. validation of the selection process
6. promotion of individual growth and self-evaluation (pp. 99-101)

In recent years there has been a movement, often based on state legislation, to give a further and important dimension to the list supplied by Bolton. It is not unusual for a statement to be made that gives a significant and possibly prime place to the improvement of student learning in teacher evaluation. A strong situation arises when there is a policy that teacher evaluation will lead not only to the improved quality of student learning, but also and connectedly to a patent improvement in the professional development and standing of teachers.

However worthy the purposes of evaluation may be, they will remain unfulfilled unless the school organization and leadership have a strong commitment to their successful completion. Lip service alone is quite insufficient. Moreover, it is essential that there is an adequate allocation of resources to plan, design, and implement teacher evaluation procedures that effectively meet their stated purposes. Stated plans and guiding principles must be supported by written assurances from those in managerial and administrative positions that all appropriate resources, including time, will be made available to attain desired conclusions.

Basic Steps

Iwanicki adopts the premise that the evaluator and teacher should share responsibility for the direction taken by the teacher to identify improvement areas in developing performances objectives. Following orientation and guidance from the evaluator, the teacher should assume major responsibility for the process that follows. This process is based on the following assumptions:

1. Self-evaluation is an essential component of the contract plan approach.
2. The evaluator is responsible for working with the teacher to develop those skills crucial to the effective self-evaluation of one's performance.
3. As the contract plan approach evolves, the role of the evaluator should become less directive and focus more on guiding, supporting, and monitoring the professional growth of the teacher in a supervisory manner.

4. Increased responsibility for the evaluation of performance should be assumed by the teacher.

Two important issues arise concerning these assumptions. First, professional growth arises from a teacher's clear discernment that it is his or her responsibility for this development. Second, because a school will have considerably more teachers than evaluators, the teacher must conduct a complete self-evaluation of performance before the initial conference with the evaluator whose time will undoubtedly be limited. In advance, the teacher must identify areas for improvement, having as honestly as possible assessed his or her performance. It is the evaluator's task to convince the teacher that the conferencing process that follows will have professionally worthwhile outcomes and become the basis for the evaluation process.

Emphasis on self-evaluation is simply a recognition that it is the teacher who has to improve, must have thoughtful ideas of ways in which the improvement may take place, and must wish to take the initiative, provided that the direction being proposed is, in the opinion of the evaluator, valid. Self-evaluation and increased responsibility of the teacher in no way diminishes the leadership role of the evaluator. The evaluator will lead as the need arises. The five basic steps follow.

1. Teacher Conducts Self-Evaluation and Identifies Areas for Improvement.

Teachers must have available to them self-evaluation techniques to identify areas for improvement. One useful approach, most likely used in conjunction with others, is to compare performance against job description. The higher the quality of the job description, the more easily it can be used as a basis to identify a teacher's strengths and weaknesses.

Next, the teacher should prioritize areas for professional development, a task made difficult and subjective, since personal judgment is involved. For this reason it is useful to enlist the aid of peer teachers and supervisory staff. Iwanicki lists the following as crucial factors in delineating areas in which change in teacher behavior is desirable:

1. time required to bring about the change
2. personnel, material, and financial resources needed to bring about the change
3. impact of the change on teacher performance in the classroom
4. impact of the change on pupil learning
5. impact of the change on the accomplishment of priority school or departmental objectives

2. Teacher Develops Draft Performance Contract(s). Iwanicki defines a performance contract as "a plan for describing, monitoring, and evaluating the profes-

sional development activities of a teacher." He goes on to propose the following basic format to develop performance contracts (1981, p. 215):

1. **Performance objective:** a statement of (a) the area needing improvement, (b) the rationale for focusing on this area, and (c) the outcome or product the teacher hopes to accomplish in this area
2. **Plan of action:** a description and schedule of activities relevant to accomplishing the performance objective
3. **Special operational requirements:** the material and personnel resources needed to accomplish the performance objective
4. **Procedures for evaluation:** the procedures to be used to evaluate progress toward and achievement of the performance objective

Such a format is flexible and designed to meet the varying needs of the school organizations and of teachers themselves. The nomenclature may also change with the term "performance contract" giving way to "improvement plan" or "professional development plan." In other words, the naming of the form and its subsections must be acceptable to all parties concerned so that the process may go forward unimpeded by disconcerting or unfamiliar wording.

A clear distinction must be made between student performance objectives and teacher performance objectives, the first relating to outcomes anticipated by students, and the second describing teacher activities to facilitate those outcomes. When a teacher performance objective is being considered, the focus is placed on both objectives. Occasions will arise when the teacher's effectiveness will be discerned by student performance.

As a separate performance contract is drafted for each area to be improved, teachers are usually unable to undertake many more than three areas during the course of a year. Clearly, the complexity of the area to be evaluated is a major consideration. Even more important is not the number of areas to be professionally assessed, but the quality of the activities undertaken.

During this process of self-evaluation and identification of improvement areas, the teacher must address questions such as the following in relation to the performance contract for each priority area:

1. Does the performance contract identify specific outcomes that can be observed or measured?
2. Does the performance contract identify the means and criteria by which the desired outcome(s) will be evaluated?
3. Does the performance contract avoid contradiction with system, building, and/or departmental objectives?

4. Is the performance contract consistent with available and anticipated resources?

The teacher must also consider an anticipated completion date for each performance and whether, realistically, selective activities will lead to heightened professional competence based upon improved student learning.

3. Teacher and Evaluator Confer to Discuss and Finalize Performance

Contract. The importance of the objective-setting conference cannot be underestimated. It is essential that the climate of the meeting is open, trusting, and constructive so that the teacher and evaluator can lay the foundation for positive future development.

The evaluator must make clear the roles to be played by both the evaluator and teacher, and he or she must allay the teacher's fears of hierarchical superiority by the evaluator/administrator. To these ends mutual professional respect—based on factors such as the acceptance by the evaluator of opinions that the teacher has proposed—should develop. The evaluator must encourage the teacher to make constructive criticisms, give assurances of confidentiality, and offer sympathetic understanding of potential problems suggested by the teacher.

Other matters to be resolved during the objective-setting conference will be the priority given to performance objectives to be addressed during the evaluation, the nature of the teaching activities related to each objective, the procedures to be undertaken to monitor progress toward the accomplishment of an objective, and the criteria to be used as a basis of judgment for such accomplishments. Before the conference can be considered to be successfully completed, there must be complete clarification by both the teacher and evaluator about these issues.

The occasion may arise where such agreement is not possible. Formal procedures may then ensue involving a representative review board. In formal procedures involving the teacher and the evaluator, third parties are selected by both to help solve the impasse. It is preferable that third party personnel are familiar both with the school context and issues related to the problem (or problems) to be resolved. Assuming that there is problem resolution during such a meeting, the teacher and evaluator then confer further to decide which performance objectives will be undertaken as well as relevant procedures and evaluation methods.

4. Monitoring Teacher Progress. Iwanicki proposes that a teacher's progress should be monitored on both a formal and an informal basis, with guidance and support offered throughout the process. A monitoring of the fulfillment of the objectives by the evaluator may entail observations of the teacher's performances. Such a formal procedure would conclude with a conference at which a written report

is furnished as a basis for a discussion on further activities. Two or three such conferences could be held during the evaluation cycle.

Written statements are not made in connection with informal conferences, which are held sufficiently often to give assistance in meeting the stated objectives. Both teacher and evaluator should feel professionally responsible for calling such informal meetings as needs arise. It is stressed, however, that the evaluator should initiate informal discussions to compliment the teacher on progress made if it is possible to do so.

The performance contract is not an inflexible document. As a result of both formal and informal conferences, performance contracts may be modified according to discerned progress. Change may also occur if objectives are found to be unrealistic. Any changes mutually agreed upon must be recorded in the written conference report that is prepared before Basic Step 5.

5. Final Evaluation of Teacher Performance. Toward the end of the evaluation cycle the teacher and evaluator confer about the extent to which performance objectives have been met. This conference is based upon procedures specified in the performance contract and is open to the opinions of both teacher and evaluator. This summative conference is also an appropriate occasion either to initiate performance objectives for the next evaluation cycle or to continue or extend those already being undertaken.

Although an individual objective may not have been accomplished perfectly, or all objectives completed, the teacher's efforts may nonetheless be construed as successful and worthy of considerable praise. Much will depend upon the complexity and difficulty of the objectives, the setting in which the teacher has carried out professional activities, and the perceived responses from students.

Other unanticipated effects may well thwart the best efforts of a very conscientious and diligent teacher. If these do occur, the evaluator should commend the teacher for fine work and suggest an extension or modification to objectives during the next evaluation cycle. Any such advice from the evaluator can be given professionally only if the teacher's progress has been consistently and thoroughly monitored by the evaluator.

Strategies for Implementation of Contract Plans

During the planning stage, before the implementation of contract plans in a school or school district, all groups affected by the process must be involved. Iwanicki suggests that this could be achieved by forming a committee comprising administrators, teachers, and representatives of teacher associations. It is preferable that problems be resolved through mediation as soon as possible.

This committee has responsibility for designing inservice as an essential adjunct to the acceptance and implementation of contract plans.

Any change is likely to produce anxiety, and the evaluation of teacher performance certainly falls into this category. To ease concerns, collaboration among concerned parties must take high priority. What is proposed should be convincing to all members of the committee and educationally and professionally sound. Effective student learning must be a dominant feature of plans.

During early discussions among those involved in contract plans, the experimental nature of the process should be emphasized. In other words, if aspects of contract plans are seen to be not feasible, then the expectation is that changes will follow. If individual concerns about the evaluation process are allayed by a general understanding that changes will be task oriented and not personally oriented, then acceptance of the process should be enhanced.

When staff have been introduced to the concept of contract plans, it is necessary that they be trained in relevant skills such as self-assessment, educational objectives and criteria for their judgment, and appropriate procedures to implement the process satisfactorily. Once again, Iwanicki emphasizes the importance of staff involvement and of allowing concerns to be raised so that incipient problems may be resolved. To achieve these purposes small group activities are more effective than large.

Conferencing skills for both the teacher and evaluator are needed. Moreover, writing skills are essential for the teacher in respect to performance contracts, and for the evaluator in respect to evaluation reports.

A Time Factor. If the task is thought to be sufficiently important, then it is likely that time will be made available. Much depends upon the priority given by the administrator to the professional development of staff through the positive evaluation techniques suggested by the process of contract plans. Nonetheless, with the best intentions in the world, time may not be available in the school for the thorough evaluation of a performance contract, or contracts, by all teachers every year. In such a situation, Iwanicki suggests that schools should implement the process on a cyclic basis so that some of the staff formally participate in the process each year. Others who participate informally will nonetheless develop performance contracts with their evaluator, conduct conferences, and reach conclusions concerning further evaluation activities. These activities should involve participation in the formal process every second or third year. It is logical that those teachers whose performance is below standard, or those who are nontenured, will participate in the formal process annually.

The Evolution of the Contract Plans Approach

Since 1981, Iwanicki's contract plans approach has been implemented in numerous settings, and much has been learned about its effectiveness. Its success in strengthening or enhancing the quality of teaching and learning in schools is affected by three factors: (a) how we think about teacher evaluation, (b) how we organize for teacher evaluation, and (c) how we conduct the teacher evaluation process. The contract plans approach has been successful in school settings where there is a need for a professional, growth-oriented evaluation process for strengthening or enhancing the quality of instruction.

It was difficult for some school boards to accept the contract plans approach during the early to mid-1980s because the more prevalent paradigm for teacher evaluation was the more inspective rating system left over from the era when schools were organized more bureaucratically. As evidence was gathered that showed that this approach was having little impact on improving instruction, school boards began to explore new approaches to teacher evaluation, such as contract plans. While the current literature on teacher development and professionalism as well as the focus on schools as learning organizations make it difficult to discount the contract plans approach to teacher evaluation, many traditional thinkers still question the legitimacy of evaluating teachers in this manner. However, more enlightened approaches to teacher evaluation will not be acceptable unless policy-makers, who often fall into the category of traditional thinkers, think these approaches will be successful.

How schools organize for teacher evaluation is critical to the contract plans approach. More successful schools are healthy school organizations (Miles, 1965) where teachers function as professionals in a climate of trust (Darling-Hammond, Wise, & Pease, 1983). Healthy school organizations are those where (a) goals are reasonably clear and well accepted by staff, (b) there is good communication, (c) staff are empowered to make decisions, and (d) staff derive a sense of fulfillment from their work. The probability of implementing the contract plans approach in healthy school contexts where teachers function as professionals is excellent, especially as compared to those schools organized along more traditional, bureaucratic lines of authority where teachers are viewed more as workers than professionals.

Even in successful schools where principals and teachers function as professionals, the contract plans approach may not have an appreciable impact on student learning. The reason for this is due largely to the fact that teacher evaluation is often implemented in isolation rather than in combination with other school improvement initiatives. Teachers may all be growing professionally in these settings, but in so many different ways that the impact of such growth on the quality of learning in the school is difficult to determine. Also, effective teacher evaluation programs that

are implemented in isolation are eventually placed on the "back burner" when the next school initiative comes along. As one teacher commented, "We gave teacher evaluation a lot of attention a few years ago when the new process came in, but now we are moving into math manipulatives."

Integrating Teacher Evaluation, Staff Development, and School

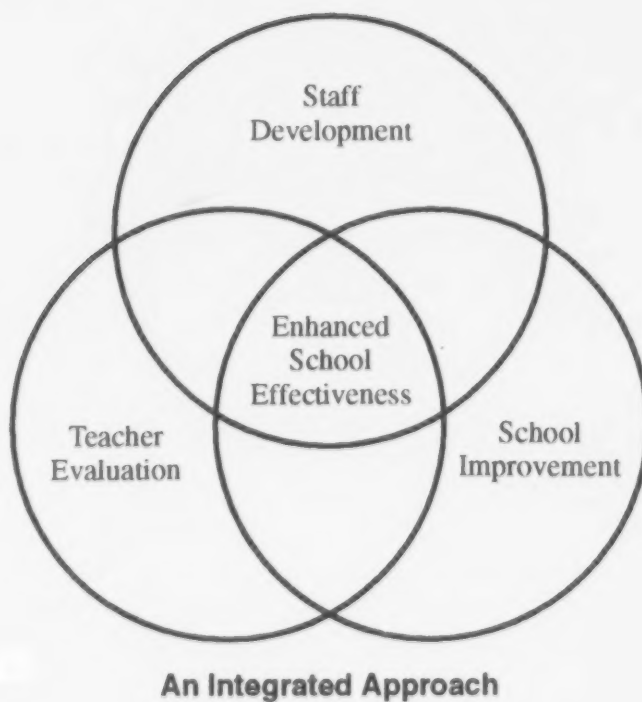
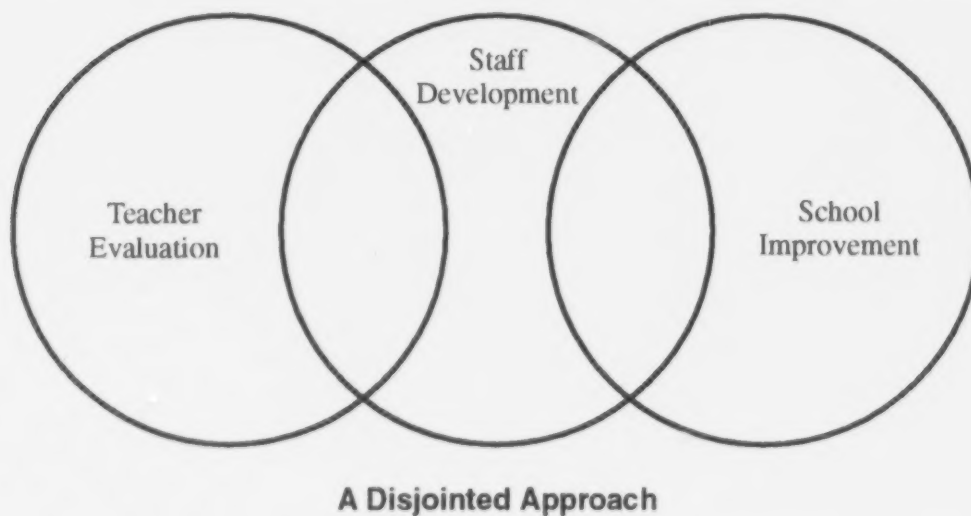
Improvement. If the contract plans approach is to enhance teaching and learning in classrooms, schools must approach the processes of teacher evaluation, staff development, and school improvement differently. In too many schools, these processes tend to be pursued in a more disjointed manner as indicated in the top half of Figure 4-1. The problem with this disjointed approach is that the limited staff development resources available to schools are often allocated independently to two processes (i.e., teacher evaluation and school improvement) that should complement each other. In fact, as noted in the integrated approach in the bottom half of Figure 4-1, teacher evaluation, staff development, and school improvement need to be viewed as three complementary processes. *The primary focus in this integrated approach is on school improvement.* Schools need to identify priority school improvement initiatives and then determine how to use the teacher evaluation and staff development processes to support these initiatives.

The processes of teacher evaluation, staff development, and school improvement not only need to be integrated, they need to be integrated through a common focus on student learning. Where do school improvement initiatives come from? They should come from what Hargraves and Fullan (1992) call "problems of practice," the critical learning needs of students that are not being met. For example, teachers in one school set the following improvement goal: *Students will meet world class standards in mathematics in three years.* Given this goal, some more specific objectives were delineated for teaching and learning as noted below.

- More problem-solving activities will be included in the teaching of mathematics.
- Students will be involved more actively in the instructional process through the use of manipulatives and group projects.
- Students will exhibit an increase in problem-solving ability on district and state performance measures.

Initially, this goal and its associated objectives set a focus for staff development. During the first year, teachers got smart about problem solving in mathematics. They addressed questions such as: What are the various approaches to teaching problem solving? How well have they worked with students similar to the ones in our school? What are some of the better measures of problem-solving ability? How does a school set performance standards? Through professional dialogue regarding

Figure 4-1. Approaches to Organizing the Teacher Evaluation, Staff Development, and School Improvement Processes (Iwanicki, 1990)



such issues, teachers developed a plan for how they would strengthen problem solving in mathematics over the next two years. The plan was framed by first identifying what students would need to know and be able to do in mathematics. Then teachers extended the plan to include instructional strategies as well as materials for teaching problem solving, procedures for monitoring student performance, and the staff development resources needed to support this initiative. It is important to note how staff development was used in two stages—first, to help teachers learn about the issues so they could develop a thoughtful plan, and second, to support teachers as they implemented that plan.

Once the plan for strengthening problem solving in mathematics was implemented, then teacher evaluation was used to support that plan. As classroom observations were conducted, attention was given to what was working well with respect to strengthening problem solving in mathematics and where additional staff development support was needed. Also, this school improvement initiative created a broad range of possibilities for teachers to pursue in developing objectives that served as the basis of their professional growth (i.e., contract plans). Some professional growth plans were even developed collaboratively by teams of teachers. As the superintendent commented, "When the process is done this way [collaboratively] there is less threat and teachers understand how it will make a difference for kids."

This example has been shared by Iwanicki to show that in settings where a continuous school improvement or total quality improvement process is in place, staff development and teacher evaluation can be used productively to support such improvement and have an appreciable impact on student learning. Since continuous school improvement is necessary for the productive implementation of the contract plans approach to teacher evaluation, it is important to build a systemwide commitment to total quality improvement through central office leadership. The *Quality Improvement Pocket Guide* (Juran Institute, 1993) describes an approach to quality improvement that is quite compatible with the direction taken in the example just shared.

As Murphy (1987, p. 160) noted, "One of the conclusions of the recent school improvement research is that schools work better when the parts fit together, when plans and activities are coordinated in a common effort to reach important school goals." By fitting the parts together through the more integrated approach, the contract plans approach to teacher evaluation has a more discernible impact on what happens in classrooms. Moreover, staff are less defensive about the teacher evaluation process, since it focuses clearly on the improvement of school programs.

Other Considerations in Implementing the Contract Plans Approach.

Generally, the contract plans approach is not the sole means by which teachers are evaluated. For example, in the Teacher Evaluation and Professional Growth Cycle

(Iwanicki, 1990; 1993), the contract plans approach is used for the professional growth component; but teachers are also evaluated on the basis of classroom observations every three to five years to ensure the public that teachers meet the school system's standards. The schedule for such evaluations based on classroom observations is not rigid. A teacher's classroom performance may be reviewed at any time if there is a good reason to do so. If as a result of such a review it is evident that the teacher does not meet the system's standards, then that teacher exits the Teacher Evaluation and Professional Growth Cycle and is placed in an Intensive Assistance Program. While in intensive assistance, the teacher is not involved in the contract plans approach to teacher evaluation. Instead, the teacher follows an Intensive Assistance Program plan designed to help that teacher meet the system's standards. If the teacher improves and meets the system's standards, then s/he leaves the Intensive Assistance Program and returns to the Teacher Evaluation and Professional Growth Cycle. If the teacher fails to meet the system's standards over time, then the school district may initiate action for dismissal.

Given the current focus on teacher professional growth and development as well as on school improvement, in time less reference is made to "contract plans." Instead, the plans developed by teachers are usually referred to as "professional growth/development or school improvement plans." As just noted, the contract plans approach is used to evaluate tenured teachers in good standing. Evaluation procedures for nontenured teachers may include professional growth plans, but the primary focus of the evaluation process is on inducting teachers properly and then determining whether such teachers meet the system's standards for instruction in the classroom.

Time is always a critical consideration in the teacher evaluation process, as mentioned earlier in this chapter. Do building administrators have enough time to manage the contract plans component of the teacher evaluation process? Iwanicki considers that they do to the extent that the teacher to evaluator ratio is in the vicinity of 30 or less to 1. There are a number of issues critical to the effective management of this approach. First, the administrator should work collaboratively with the teacher to develop meaningful and challenging professional growth plans. On the average, such plans tend to include two objectives that are attained over a two-year period. Once these professional growth plans are developed, supervisors or peer teachers, rather than principals, should support and monitor teachers' efforts to achieve the objectives included in the plans. Usually principals do not have the time to support and monitor all of their teachers' professional growth initiatives, but they need to be prepared to step in if a problem arises. Given the nature of most professional growth plans, supervisors and peer teachers are most qualified to support and monitor their colleagues' progress.

Finally, teachers should be allowed to complete the final evaluation reports on whether the objectives included in their professional growth plans have been

achieved, since their principals do not have the time to prepare them. Teachers are very professional and do an excellent job in completing such reports once the process is explained to them. Since this final evaluation report is cosigned by the building administrator before it is placed in the teacher's personnel file, the principal has the option of attaching a dissenting opinion. In schools where this practice of allowing teachers to complete their final evaluation reports has been implemented, the need for principals to attach a dissenting opinion has been negligible.

Summary and Conclusion

The effective implementation of contract plans depends upon a commitment by the school organization and all involved in the process to ensure that it will succeed. The teacher evaluation process must be consistent with the unique needs of the school as well as those of the teachers being evaluated. There must be collaboration with all involved in the process before a final design is constructed for implementation.

There are no absolute strengths or weaknesses to the process of contract plans. Crucially important, however, is the philosophic stance adopted by professional staff and the manner in which plans lead to design and implementation. If attitudes and procedures are sound, it is likely that potential weaknesses will be obviated. These include the inability of the process to rank teachers, placing too much emphasis on the attainment of measurable objectives, the overwhelming amount of paperwork, insufficient time, and having unqualified decisions reached by an evaluator who does not seek professional support and opinions of senior staff who are more qualified to judge a particular performance objective.

If, however, the process is made professionally strong from its inception, a teacher's professional growth will be enhanced, there will be a good working relationship between teacher and evaluator, teacher competencies will be brought into sharper focus to the benefit of student learning, and there will be an integration of the teacher's foremost objectives with the school's goals.

References

- Darling-Hammond, L., Wise, A. E., & Pease, S. R. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research*, 53(3), 285-328.

- Hargreaves, A., & Fullan, M. G. (1992). *Understanding teacher development*. New York: Teachers College Press.
- Iwanicki, E. F. (1981). Contract plans: A professional growth-oriented approach to evaluating teacher performance. In J. Millman, Ed., *Handbook of teacher evaluation*. Beverly Hills, CA: Sage.
- Iwanicki, E. F. (1990). Teacher evaluation for school improvement. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers*, (pp. 158-171). Newbury Park, CA: Sage.
- Iwanicki, E. F. (1993). *Teacher evaluation and professional growth in more productive schools*. Storrs, CT: The Connecticut Institute for Personnel Evaluation, Department of Educational Leadership, The University of Connecticut.
- Juran Institute. (1993). *Quality improvement pocket guide*. Wilton, CT: Author.
- Miles, M. B. (1965). Planned change and organizational health: Figure and ground. In *Change Processes in Public Schools*, (pp. 11-34). Eugene, OR: Center for the Advanced Study of Educational Administration, University of Oregon.
- Millman, J. (Ed.). (1981). *Handbook of teacher evaluation*. Beverly Hills, CA: Sage.
- Murphy, J. (1987). Teacher evaluation: A conceptual framework for supervisors. *Journal of Personnel Evaluation in Education*, 1(2), 157-180.

Getting Value from Teacher Self-Evaluation

By Graeme Withers

Introduction

Graeme Withers, of the Australian Council for Educational Research, emphasizes that evaluation has importance in the daily lives of teachers. He refers to self-appraisal as a professional duty that benefits both teacher and learner. He argues that self-appraisal can and should be held to rigorous standards of teaching performance and student progress and need not be self-serving. He says that such self-appraisal should be ongoing and should provide the basis for planning annual teaching programs based on what worked best in the past. This becomes part of a continuous formative approach based on "a structural, integrated, but holistic evaluation" of the evolving performance of both teacher and class.

Withers broadens self-evaluation to "co-professional evaluation," evaluations by colleagues of each other's work and against criteria of sound teaching and student progress. He argues that teacher-generated evidence and external assessments by a co-professional are potentially richer and hence more valuable than either source of evidence alone.

Withers projects that regular use of self-appraisal may well increase a nation's chances for improved teaching and student learning. He also says that effective self-appraisal and appraisal by co-professionals could provide a basis for holding off external, mechanistic evaluations of teachers by demonstrating that the profession appraises and evaluates its performance from within.

Professional Autonomy

To put one's arm around a stressed-out colleague who has just reeled into the staff room after a particularly bad Friday or to make her a cup of coffee and feed her the cake left over from morning tea are probably necessary responses from a co-professional seeing another human being in need of support. But are they sufficient responses to the larger predicament of the physical and intellectual tensions that beset the professional lives of teachers?

Here's another response. At an English school that was very new when I visited it a few years ago (Greendown School, in Berkshire), the foundation principal had established just two working principles for the development of the school's program. One was "no bells or public address system of any kind." The second was more radical: "no teacher is to be in a classroom with students for more than 20 minutes without the presence of another staff member."

What price the professional autonomy of the classroom teacher in the circumstances of the latter principle? Such autonomy—a high degree of freedom to plan and conduct one's own program in one's "own" classroom—is greatly prized by lots of Australian teachers and jealously regarded by many overseas teachers. Those elsewhere whose professionalism is constrained by very different organizational and "management" principles, such as test-based or learning-kit programs that determine their everyday practice, look longingly at our classrooms. But does such autonomy work, and is it the best way of working? These are large questions, not capable of easy answers but very susceptible to glibly negative ones.

Challenges From "Above"

What the Debate is About. There are signs that, Australiawide, the debate about autonomy is really hotting up. An interesting complex of issues is emerging currently and simultaneously joining a few hoary chestnuts, like the "whole language versus genre" side taking, which already divide professional opinion.

For example, one of the issues that divides the profession (and may divide it even more sharply in the near future) is the challenge from the growing demands for centralized, even national, curricula—external specification of exactly what it is teachers should actually be doing in the content of their programs. This is linked clearly in the current rhetoric to achievement or attainment "targets" to be met by students (and hence, in a sense, by their teachers) when they learn that content. The challenge to professional autonomy from overarching systems, as rule givers about content, process, and the standards that might be expected to result is considerable.

A second (linked? related?) strain comes from current calls for some system of teacher appraisal—what some regard as the "New Inspection." This, too, divides the profession quite sharply. The views expressed at meetings of senior administrators and principals are not likely to be those aired at formal or informal meetings of rank-and-file teachers.

Autonomy and Evaluation. At the same time it is a debate about autonomy and professional freedom, this is, of course, a debate about evaluation. Bloom and others long ago told us what evaluation meant in learning terms: they distinguished between judgments based on internal evidence and judgements based on external criteria. They also said:

Evaluation is defined as the making of judgements about the value, for some purpose, of ideas, works, solutions, methods, material, etc. It involves the use of criteria as well as standards for appraising the extent to which particulars are accurate, effective, economical, or satisfying. The judgements may be quantitative or qualitative, and the criteria may be either those developed by the student or those which are given to him (sic). (Bloom et al., 1956, p. 185).

For "student" in the last sentence, one might also read "teacher" if it is her work to be evaluated.

National (or state) attainment targets and teacher appraisal are each intended to achieve an evaluation: an external view of the health of the various education systems through looking at their processes, and the performance of students who learn in the schools. The current understanding of the means by which such large-scale system evaluations would be best done does not seem anywhere to include direct evidence from within the rooms in which the learning takes place.

The Argument

My argument here is that, on several counts, the view from the teacher's desk ought to receive as much attention as public and political perceptions of the importance of national curricula and the need for teacher appraisal. Teachers are (or ought to be) practiced evaluators: they can contribute much. They will also be the principal subject of such evaluations: we risk missing out on the very intention of such evaluations—improvement of the status quo as regards standards and practice—without their willing compliance. Furthermore, given a decade or so of teacher bashing, the attractions of the profession must be supplemented rather than diminished. The attractiveness of the profession to the best possible level of entrant needs to be enhanced rather than merely maintained. And chief among these attractions needs to be professional power.

In addition, any new appraisal "systems" must use techniques of evaluation that have been proven to be effective. These ought to include the notion of monitoring progress over long periods, rather than the "big bite" approach. It would be possible to use frameworks and even criteria not unlike the ones teachers apply to their students. The design of such systems would need to include commissioned professional contributions from rank-and-file teachers—a start to collecting the internal evidence that Bloom talks about. And I don't mean gestures, as in the Victorian Literacy Profiles development, but real contributions at all stages of development. The large-scale planning of appraisal strategies will otherwise be deficient, and their administration is likely to be counterproductive in the worst imaginable ways—even lower morale, even more stress, an increased flight from the profession.

A Notion of "Co-Professionalism." In coming to terms with these matters, teachers themselves might have to redefine or adjust their concept of just what professional autonomy ought to be. The notion of the primary classroom teacher as queen of her castle might need to take a real battering and be replaced by some greater degree of what, for want of a better word, I'll call "co-professionalism." The predominant school organizational mode in primary education still seems to be nuclear classrooms, with the teachers somewhat isolated within them. Schools already make inroads on that isolation, but rarely on any consistent or sustained basis, like Greendown School. And they are unlikely to in the future. In-school professional development, occasions for school policy development on specific issues, external in-service education, and (particularly) support- and team-teaching of various kinds go some way to reducing the isolation. But one wonders if these are enough.

Let's learn from programs such as the Early Literacy In-service Course (ELIC) and all those other clone programs of teacher development. There seem to me to

be two main reasons for the success and impact ELIC has had in the United States, for example. One is that programs like this feed the local teachers' natural desires for greater professional power after generations of instruction by tests and learning kits. A second might be the convincing demonstration such programs provide that learning at any level—professional or student—is not all individual self-study. It is more individually effective when it is situationally cooperative, as in an ELIC sharing session—and this cooperation looks as if it is the crucial element.

Here's Kate, an Australian teacher of eight-year old children, giving a simple example of the cooperation and where she saw it had led to, in terms of her classroom work and her students' achievement:

Over the last twelve months, we as a staff have really looked at how effectively we have catered for individual needs. We realized we were satisfied with the content of what we were teaching, but it was very much the teacher who had power in the classroom—teachers were planning the activities, and how they would be taught. We questioned to what extent we were empowering the children to reach their full potential, and to become more in control of what they learned. As a result we began to look at children negotiating the curriculum. After a great deal of background reading, and many discussions and in-services, we began to negotiate with the children. Children know how they learn best, so I had to let the children in my class have enough freedom to negotiate their activities and how they would complete them.

Through negotiating, I watched all the children in my class develop an even greater enthusiasm for learning, and have learned a great deal myself from listening to their ideas. Negotiating caters for the wide variety of individual needs that exists within my class. By negotiating their own ideas, a class topic can see 24 very different pieces of work being presented. This eliminates comparisons between children, as everyone's work is seen as original, and it takes the pressure off children who find learning difficult.

Within my class the children work within very clear guidelines, but at the same time [they] know that there is a definite place for their own ideas. Children who find learning easy can really extend their thinking through negotiating and challenge themselves to achieve their best. Children learn best when they are interested in what they're doing, and I've found that negotiating certainly creates interest. . . . Negotiating is not a simple process to implement in any classroom, but I feel I've at least made a start, and achieved very favorable results.

Let's also learn from two related research studies: one in Scotland, the other local. Brown and McIntyre (1989) and Batten, Marland, and Khamis (1993) have been working with teachers on projects exploring what both studies call the professional craft knowledge of teachers. They maintain that much such knowledge (from impromptu but reasonable responses to classroom situational demands through to deep understandings about valid and powerful strategies for teaching and learning) often largely remains tacit—perhaps unrecognized by the practitio-

ner—until articulated in company with their peer teachers. Once articulated, it is there, patent rather than tacit, for all to share and develop further. Where is the best craft knowledge gained? In the craft place; that is, the school (rather than the nuclear room), but only if one talks about one's own with co-professionals and lets a critical light shine on it.

Those insights might constitute a means of adding a measure of self-appraisal, self-evaluation, to the professional business of being a teacher. But self-knowledge about one's craft is obtained and worked on and developed in a co-professional situation by actively using the insights and experiences of others to both illuminate and extend the teacher's view of herself, as well as taking the opportunity to explain and defend what it is she already does or knows, whether consciously or not. If done sympathetically, in congenial learning situations, it will not diminish her essential autonomy but inform and improve her day-to-day practice.

It would need to be done formally, and at all levels of expertise, not just beginner teachers. Not everyone has the time and inclination to do an in-service course or a graduate diploma—the school itself functions as a site for the activity. It could start anywhere. Who are the fellow teachers we admire? The ones whose rooms buzz, but where stress is replaced by a working harmony? Might they not have a lot to tell us about how it got to be that way? How do we get them to tell us?

Here's an example. Just recently (1993) I worked with teachers in three high schools: each school determined that it needed to spend time investigating how strategies for better teaching and learning might be developed. Each school established a different theme, or need: one school wanted to implement a Language Across the Curriculum policy; another had a problem with its 11-12 year-old intake; the third wanted to improve practice generally across the school.

In each case, even given the different themes or slants, the following program was conducted. The entire staff met for a full day: an introductory session allowed five teachers to make a presentation each, outlining or role playing one prized strategy from their classroom that they thought related to, or exemplified, the theme. Then a brainstorm session occurred—each teacher sat down with a pack of blank library-system cards and on separate cards briefly recorded as many tried, proven strategies from his or her experience that constituted "good practice" for that individual. During lunch time, those cards were sorted and when the staff met again after the break, they discussed, in groups, a small selection of the morning's cards that exemplified some larger policy or practical issue related to the day's theme. The group was asked to develop a statement of policy on its issue and support it with practical strategies taken from the cards or a further group brainstorm. In the school that had a problem with its 11-12 year-old intake, these group issues included "independent learning," "teaching in mixed-ability classrooms," and "teaching those unlike ourselves."

The process didn't end there. The cards were taken away, word processing ensued, and the resulting list of individual strategies were sorted under headings—"Classroom Ethos and Atmosphere," "Setting Up Conditions for Creativity," "Students as Teachers and Models," and any others that had developed naturally as focal issues for the teachers concerned. The group summaries together with the sorted strategy listing, once indexed (because some rich strategies had relevance to more than one heading), then became the text of a school handbook, and a copy went to every teacher. That handbook represented the accrued wisdom of about 600 years of teaching experience—a total achieved when one adds up the individual years of service to the profession given by all the staff.

Even then, the process was not complete—what was each teacher to do with the document in terms of enhancing his or her practice? A specially-written introduction made it clear that the book could operate on at least three levels:

- as an opportunity to stimulate individual reflection by each teacher about one's own or other's practice
- as a resource bank for groups of teachers—either those within a subject teaching department, or all those who teach at one year level, or those who share a common interest in, say, "concept development and refinement," to quote one area of interest in one school
- as a foundation for the development of schoolwide consensus and policy about improved practice in any area of perceived need

The key word emphasized in that introduction was "translation." What works at one year level or in one subject department such as mathematics or art, may very well work in another subject area or at a different year level. The assembly and layout of the handbook were intended to promote, individually and in groups, such "translations" of practice. Do the schools use the books? Yes, they do. How are they used? By teachers coming to the policy statements and strategy lists with their own programs in hand, and comparing current and possible practice—making new choices about what works best.

Getting Practical

Heightened professional knowledge about what constitutes superior teaching both contributes to improvement in classroom practice and feeds off such practice. So, in this next section, I want to focus on one aspect of practice—program planning—and look at its implications in order to lay down a few rules (er, sorry, "propose a few guidelines") for both self-evaluation, which is the duty of the professional, and

eventual improvement of performance in both student and teacher terms, which is the aim of each professional.

One is wary of suggesting anything that places greater strain on teachers' available time, so the trick might be to get maximum benefit from knowledge common among, and efforts commonly made by, today's practitioners. As noted above, useful and valuable strategies for the monitoring and evaluation of student progress can be used for monitoring and evaluating teacher practice as well.

An Aspect of Practice—Program Planning. One might start by asking: "How does a teacher's program for her year with a class ever get planned?" On her desk sit departmental guidelines, frameworks, and syllabi, together with whatever other documentation the school can contribute. In her head lie last year's experiences, together with a vast range of ideas and strategies that form part of her general professional preparedness. She puts all these together as best she can.

The specifically evaluative aspects of program planning might now emerge. What are some of the key features of the best programs against which her program might be compared? Leaving questions of content aside, teachers with whom I have worked in various research studies point to these, among others:

- planning in substantial periods of time so that students have the opportunities to work through to a finish they can feel satisfied with, and proud of
- planning for sequences of instruction (themes, topics) rather than whole terms or semesters, because these are too long to allow for necessary changes
- using headings rather than microscopically detailed work plans so that some spontaneity is preserved
- building in possibilities for individual students to interpret set objectives and desired learning outcomes, and to add personal goals in addition to teacher-set objectives
- developing a dynamic and varied set of learning and teaching strategies within whatever structured formats the teacher might choose
- aiming at maximizing individual progress and development rather than merely uniform maintenance of externally-ordained "standards"
- relating to conditions and life outside the classroom—the general reality within which the students live

An Implication of Practice—Teacher Self-Evaluation. Once the program is designed, the teacher-planner needs to ask herself a number of broad questions, even before that program gets to be implemented in the classroom.

One might be: What really works for me in my classroom—what procedures for teaching, learning, and assessment? Subsidiary questions then emerge: What do I do well now? What not so well? What might I contemplate doing to improve my

performance, based on the experience of others? How might I develop beyond my current level of professional understanding, so that overall this program runs better than last year's?

Another broad question will duly emerge, to be considered in detail once the program is under way. How will I know it's really working, in both my terms and the students' terms? One then needs to become very conscious of specific aspects of the program's delivery: What strategies will I have to put into place for monitoring any improvement in student learning outcomes over my expectations or their history? How will I develop the necessary criteria for judgment? How might I monitor my own reactions, both physiological (tiredness) and psychological (stress), to the way this program operates?

In general, these are reflective processes—they will have maximum value when they are undertaken not just as gasps at the end of a working day but consciously, definitely, and regularly, in a planned approach.

Another Implication of Practice—Co-Professional Evaluation. From the complexity of these questions, it might be clear that few teachers will be able to enter into such an evaluative process without help. Also, working from the premise that evaluation of one's own and the students' performance will (or ought to) contribute to improvement of this year's classroom practice, as well as spin-off for one's general professional development, it becomes obvious that the evaluation, like the program, will have to be ongoing.

For a program, however it is shaped, to work (in Bloomian terms) effectively, economically, and in a way satisfying to teacher and students, I believe it needs, for a start, to be in some way (or at some level) accredited by co-professionals. One needs at least to sit down with a colleague and jointly review the programs each has designed according to a set of criteria such as the one offered about "good" programs in a previous section of this article.

Such joint evaluations will certainly involve student assessments, and perhaps even a few measurements of progress. A little cross-assessment of the products of learning from one another's classrooms might take place. Observations of one another's teaching will add insights for more reflective conversations about practice generally. As I've suggested above, the evaluation, like the program, will need to be ongoing, and the reflections regular. Like all learning, this learning about self will accrue gradually with occasional leaps, and it might need to be captured in a little detail—keeping a professional journal or log, which at least notes the leaps, the key insights from the reflective conversations. A reviewer, looking at an early draft of this article, made the comment:

Teachers also need to know how to critique each other's work and be critiqued. Otherwise this won't do much good. They need training in this.

I have to agree. However I'd contend that activities such as those I am suggesting in this article at least familiarize teachers with the processes—make them conversant with and used to professional scrutiny, and sensitized to what that means for themselves and others. Teachers, after all, are often loath to do such things. And one of the less desirable aspects of the organization of schools into "nuclear classrooms" is that time constraints become all-powerful: teachers are on duty all the time. Often, too, feelings of personal inadequacy (not just lack of training) get in the way. Defensiveness, rather than assertiveness, becomes the leading mode. But many teachers feel threatened simply because they don't recognize how good they are. They don't give anyone else an opportunity to tell them: the daily stress causes them to ignore their long-term, overall strengths, and it reduces their actual successes to the status of mere "accidents," days when things just happened to go well. Sometimes, too, a somewhat misguided view of professional autonomy means they feel under threat in a different way—that their "right" to conduct their own program in their own ways is being challenged. As a profession, they will certainly have to struggle to find practical ways around these difficulties of time, attitude, and lack of training. To do so may well be vital.

The Message From "Above"

The message seems quite clear to me. Teachers need to up their game in the matters of expertise and practice in self-appraisal and self-evaluation and link these procedures to co-professional accreditation of their individual professional efforts at the school level. If they don't, then Somebody Up There is going to do the appraisal for them, externally and mechanistically, rather less harmoniously and more threateningly. In the event of such interventions, what one might call the real professional autonomy may well disappear. So far the specter of national testing has been held at bay, in Australia at least—but I doubt the profession's chances of holding off external appraisal procedures similarly, unless teachers can demonstrate that the profession appraises and evaluates its performance from within. For that to work soundly, I believe teachers need more practice at it than they give themselves at the moment.

For any of Those Above who also chance to read this chapter, I would want to add two riders to the discussion, opinions rather than facts like much of the rest of this chapter. One is that to cede any degree of autonomy to individual professionals is not to diminish (or seek to diminish) the acceptance and applicability of standards, either of professional behavior or student performance, whatever they may be. It may well be to increase the nation's chances of actually achieving improvements in both. The second is that self-appraisal is not necessarily self-seeking. On the contrary: I suspect that teachers, like students, are in fact inclined to be

overly critical of their performance when they evaluate themselves and that performance.

Summary

In this chapter I have tried to focus on ways and means of evaluating a teaching program, as distinct from making assessments or measurements of student achievement. However, experience and good practice in the latter will obviously contribute to the former.

I have also tried to focus on means of achieving a structured, integrated, but holistic, evaluation of the performance of teacher and class throughout a year, not a piecemeal view of bits of the program but rather the whole as it develops.

I have attempted to demonstrate a belief that evaluations conducted using internal evidence (from the person being evaluated) and external views (from a co-professional referee) are potentially richer and hence more valuable than those conducted by one party only.

I have also suggested that for such an evaluation to be most effective in promoting learning and raising professional expertise simultaneously, it will need to be a continuously formative evaluation—drawing strength from achieved successes and designed to contribute to further improvement.

References

- Batten, M., Marland P., & Khamis, M. (1993). *Knowing how to teach well*. Melbourne: Australian Council for Educational Research.
- Bloom, B., et al. (1956). *Taxonomy of educational objectives: Handbook 1: Cognitive domain*. London: Longmans.
- Brown, S., & McIntyre, D. (1989). *Making sense of teaching*. Edinburgh: Scottish Council for Research in Education.

Richard Manatt: Teacher Performance Evaluation

One who has addressed the growing concern of school districts and the public about the need for improved teacher performance is Richard Manatt, professor of education and director of the School Improvement Model (SIM) for the Research Institute for Studies in Education, Iowa State University. During the late 1970s he accepted and developed the teacher performance evaluation (TPE) approach as a model for

teacher evaluation and development. He considered TPE to have a sound theoretical and philosophical base.

To promote the concept of TPE, during the 1980s he developed videotapes and accompanying materials for use during seminars and workshops. These activities have resulted in large numbers of administrators and senior educational personnel being strongly influenced by Manatt's cogent and convincing approach to teacher evaluation. Numerous school districts, particularly in Iowa, have adopted TPE as an effective model for assessing teachers and developing their competency.

Commenced in 1979 and concluded in 1983, the very impressive School Improvement Model Project has placed TPE in a context of total school improvement, thus enhancing its importance. The School Improvement Model Project, a very significant undertaking involving two school districts and one independent school in Minnesota and one school district in Iowa, investigated the effects of a systemwide (or schoolwide) articulated system of administrator and teacher performance appraisal on student achievement. The very real benefits of the outcomes of this study have become important components in national school/teacher effectiveness workshops organized by Manatt and a Codirector of Iowa state University's SIM projects, Dr. Shirley Stow.

Although this chapter will dwell on TPE, the complete picture of Manatt's contribution to the practice of teacher evaluation demands reference to the SIM Project.

Introduction

Teacher performance evaluation is based upon an analysis of measurement of progress made toward the accomplishment of predetermined objectives or, as Manatt calls them, job targets. It does not follow the line of traditional product-process approaches (or input/output) but is based upon a process that depends strongly for its success on an understanding by both teacher and evaluator of what constitutes effective classroom instruction. It also insists upon effective and efficient use of time. In a leader's guide, accompanying a videotape for staff development, Manatt (1982, p. 3) stated that to be successful, TPE requires:

1. Rating scales with criteria based on effective teaching research
2. Lesson analysis in conjunction with skillful observation
3. Coaching and counseling techniques that motivate teachers to change
4. Provision for procedural and substantive due process of law to provide protection for both teachers and evaluators

Philosophical and Theoretical Bases for SPT. In an endeavor to move away from an input/output view of teacher evaluation, Manatt accepted the criteria proposed by Strike and Bull (1981) for teacher evaluations that were both legal and morally acceptable. The school system must

1. Make clear the formal administrative policies of the school board and provide a reasonably precise explanation of the criteria aimed at assuring both effective teaching and uniform procedures for making personnel decisions
2. Guarantee that the evaluation will focus only on those aspects of a teacher's performance, behavior, and activities that are directly or indirectly relevant to the teacher's ability to execute the legitimate responsibilities of the job
3. Allow teachers legal due process
4. Develop nonalienating, productive, and cooperative working relationships and aim at increasing the professional skills of teaching staff
5. Share the evaluation data with teachers and provide necessary assistance

Manatt found that the Teacher Performance Evaluation (TPE) model incorporated these criteria into a coherent system that he further refined by giving a cyclical emphasis to the process. By so doing he showed that the major purpose of TPE is the improvement of instruction. Moving away from the input/output approach to evaluation, he has improved the image of teacher evaluation by showing that the purpose is not to weed out poor teachers, but to upgrade the competence of all.

Although his TPE model has both formative and summative aspects as part of the process, the latter is viewed more as a mechanism for improvement than as an instrument to dismiss poor teachers. This aspect of the process will be examined in further detail later in the chapter.

Differentiation Between Teacher Performance Evaluation and Clinical

Supervision. There has been considerable confusion over similarities and differences between teacher performance evaluation and clinical supervision. The significant difference is that teacher performance evaluation goes beyond clinical supervision to record accomplishments for future decisions about a teacher's classroom development.

TPE also differs from clinical supervision in that it analyzes how teachers are giving instruction by calling the teacher's attention to the organization's requirements and also to student achievement data. It aims to build in quality control mechanisms. While it is somewhat similar to clinical supervision in that it gives emphasis to the classroom curriculum, it differs in that it compares one teacher's performance against that of another.

There is further divergence between teacher evaluation and teacher supervision when it is considered that evaluation causes school organizations to make plans and specifications regarding criteria for effective teacher performance. In addition, a district will be monitored against those standards and an appropriate reporting mechanism will be instituted. Clinical supervision, more so than teacher evaluation, requires all staff members to identify their individual strengths and weaknesses.

To sum up, the significant difference between the two processes is that teacher performance evaluation is based on analysis and measurement of the progress the teacher makes toward the accomplishment of predetermined objectives according to policies formulated by the school district or school. Clinical supervision is based on teacher instructional improvement by a professional monitoring process.

Introducing Teacher Performing Evaluation to a School or School District.

Manatt then stresses that the introduction of TPE to a school district should not be a rushed process. He suggests that three years is an optimum period for planning, although this may be shortened depending upon the circumstances, such as the time taken for all involved personnel to make decisions about what plans or approaches are to be followed, and to accept these. One section of this chapter will deal with the establishment of a steering committee and rule-setting procedures for the development of a personnel evaluation system.

Another part of the chapter will look at the actual steps in the TPE cycle. Considerable emphasis will also be given to guidelines for classroom observation including criteria for satisfactory levels of teacher performance.

Although space will allow only the more salient features of TPE to be given, its placement within the full context of the School Improvement Model must be recognized. The research carried out by Manatt and the team from Iowa State University on the SIM project has drawn the conclusion that the evaluation for teacher improvement, to be effective, must be seen in the context of districtwide or schoolwide commitment.

Teacher Performance Evaluation in the Context of the School Improvement Model

If the baseline goal of a teacher evaluation system is the improvement of student achievement, successful outcomes are likely to be achieved only if the teacher evaluation process is linked to other important components of the school system. The school improvement model (SIM) has been a demonstrated way of improving student achievement by way of improved teacher performance as a significant factor.

SIM involves all major aspects of a school system as it pursues its stated aim of raising student learning and achievement levels from K-12. SIM endeavors to make four important linkages. In link one, teacher performance is described, evaluated, and related to student learning; in link two, administrator behavior is related to teacher performance; in link three, there is the requirement that the functional classroom curriculum—course content as well as instructional methodology—and the testing techniques match the goals and aims of the school community; in link four, interventions in the form of training, changes in instructional strategies, and improvement of leadership are created.

All four linkages directly and indirectly require that both teachers and administrators are evaluated to determine the extent to which they have made progress toward the accomplishment of predetermined objectives. Under the SIM approach administrators are assessed under the process known as Administrator Performance Evaluation. Teachers are assessed using the TPE process, which analyzes and rates teacher performance on a wide range of criteria that are **valid, reliable, and legally discriminating** (that is, capable of explicating differences between productive and unproductive teachers). Whether TPE is used under the umbrella of the SIM process or apart from it, Manatt maintains that there are four fundamental questions that teacher evaluation, to be effective, must address:

1. What are the criteria of the desired teacher performance?
2. How high are the standards that the district wishes to set?
3. How will the district monitor, report, and measure a teacher's progress?
4. How does the administrator help the teacher improve?

It is essential that all school district personnel involved with TPE understand its purpose and rationale. The many activities associated with TPE, aimed at determining the level of a teacher's performance and improving the quality of the educational program, make it imperative that careful planning tailors the evaluation system to fit the needs of the particular school district. There must be prior approval and support from the school board, there must be board representation in the planning process, and all that is proposed must be congruent with the district's instructional goals and philosophy.

It should be recognized from the outset that whereas the school board has the legal right to determine evaluation criteria—what is to be achieved and the desired levels of achievement—the actual evaluation procedures are negotiable under the bargaining agreement. These facts must be clearly understood by the steering committee.

Developing a Performance Evaluation System

Although the focus of the remainder of the chapter will be placed on teacher evaluation, it should be understood that the other three linkages of the SIM approach are assumed to be occurring. For instance, Figure 4-2, which is a flow chart for developing a performance evaluation system, could be construed to include administrator evaluation as well as teacher evaluation. Figure 1 indicates that the iteration from the presentation of a proposal to the board, to the selection of subcommittees, to field trials and revisions, and to the implementation of the system could take up to three years. As the flow chart is generally self-explanatory, only those aspects needing clarification will be discussed.

The Steering Committee and Its Subcommittees. The steering committee, representative of teachers from various departments and grade levels, administrators, board members, community members, and students from the secondary level is selected and organized. Numbering no more than 20, its prime function is to guide the development of the system. Its major planning tasks include identifying needs of the district, particularly in respect to student achievement; determining the scope, sequence, and time line of the evaluation system; and communicating with the board for decision-making purposes and with the superintendent and staff for consultation and information reasons.

The steering committee must address the same four key questions mentioned in the previous section, the most crucial being: What are the effective criteria of effective teaching?

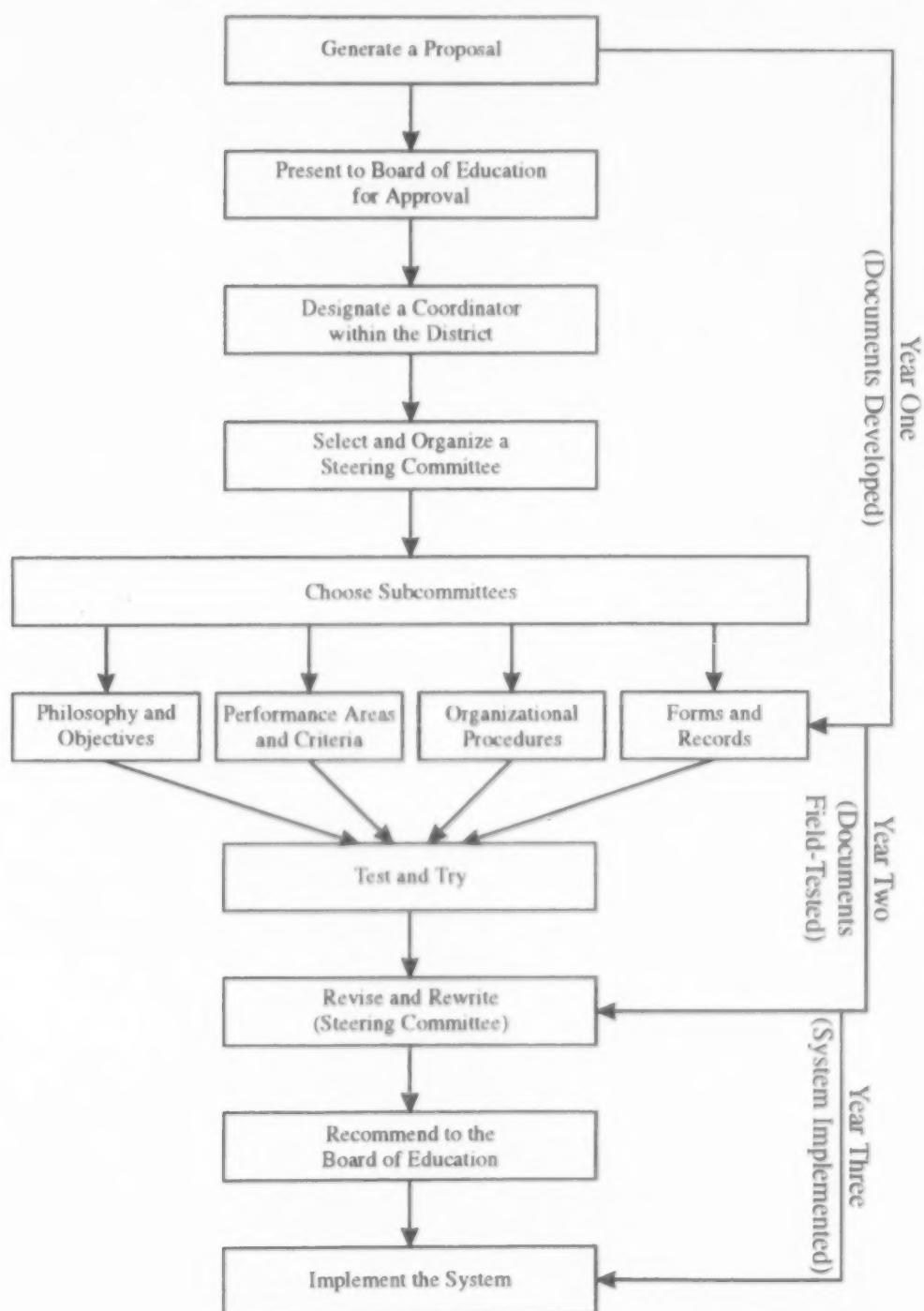
Five subcommittees propose and present solutions to questions assigned to them. To the extent possible they set specifications for the system, at least on a trial basis, by carrying out assigned tasks:

- to define what good instruction and effective administration means in the district
- to define the reasons for evaluating teachers (and other personnel under the SIM approach)
- to decide major responsibilities of various administrative personnel
- to decide how many evaluators to use

Performance Areas and Criteria Subcommittee Responsibilities:

- to determine the performance areas to be considered
- to decide what specific areas to include in the evaluation
- to define the specific criteria to use

Figure 4-2. Developing a Performance Evaluation System



Operational Procedures Subcommittee Responsibilities:

- to establish how to use multiple evaluators
- to decide what the cycle should be, what an observation is, and how to give feedback and help
- to determine who should handle the appraisal interview

Forms and Records Subcommittee Responsibilities:

- to analyze the system's paperwork and documents
- to determine the need for different documents for observing and reporting
- to define work samples

Test and Try Subcommittee Responsibilities:

- to determine an appropriate test of the system
- to determine validity, reliability, and discriminating power of the criteria and to recommend starting time of field tests
- to define orientation and training of the evaluators
- to recommend modification of the system before the formal adoption

Activities Generated by the Various Subcommittees. As the subcommittees meet they generate ideas that are further discussed with the steering committee for tentative decisions to be made. Subcommittee decisions are then used to develop activities, procedures, and prototype instruments.

For example, in relation to the teacher performance evaluation system decisions have to be made about philosophies, performance areas and criteria, the evaluation cycle, and job improvement targets. While all these areas will be subject to trial, further discussion, and revision, initial recommendations are nonetheless important. Thus, philosophies of educational instruction and evaluation are developed, bearing in mind the needs of the school or school district and its particular culture. Performance areas, within categories such as productive teaching techniques, classroom management, and less organization, are identified. Standards are usually set by the school board and administration (for purposes already given), but it is the steering committee's task to establish the procedures for the evaluation cycle. As we shall see later, in Manatt's TPE model both formative and summative types of evaluation are included.

The steering committee, in conjunction with its subcommittees, establishes such procedures as length and frequency of classroom observations, who is involved in evaluation as evaluators and evaluatees, and due process considerations.

Four documents that monitor and measure teacher performance are developed for trial purposes.

1. The preobservation data sheet contains the framework for classroom observation and for the preobservation conference. It may include such items as the objectives of the lesson, teaching procedures to be used, special characteristics of students to be noted, and specific teaching behaviors to be observed. The document prepares both the evaluator and teacher for the classroom observation that follows.
2. The formative evaluation report will contain the information obtained during observation(s). In effect, the form is a working document.
3. The summative evaluation report records progress made toward achieving the specified objectives and standard expectations for all teachers in a particular school or school district. The judgments made on this report are based upon both formal and informal reports and data gathered in connection with the formative evaluation report.
4. The Job Improvement Target form is used to record between three and five teaching goals (or job targets) agreed to by teacher and evaluator. Specific objectives are stated, which the teacher will endeavor to attain. The form will also set a time limit for reaching a target and prescribes measurable ways to determine the extent to which it has been reached. The form enables teachers and evaluators to focus on areas where improvement is needed and to set goals for such improvement. This evaluation occurs at the conclusion of the time period specified in the job target.

Field Testing. After the above components for the performance evaluation system have been developed, the prototype instruments are field tested. The try and test subcommittee develops plans for field testing, for the orientation and training needed for implementation, and for monitoring the appropriateness of forms. It is assumed that changes can be made easily in response to unanticipated effects. During field testing, assistance and responses are sought from parents, students, teachers, and administrators.

Following the field testing, data are reviewed and used to refine the system. Recommendations for implementation are then made to the board, usually during the third year of the developmental cycle. In general terms, documents are developed during the first year in response to the work of the subcommittees, field testing is carried out in the second year, and implementation of the TPE system is then possible toward the end of third year.

The TPE cycle stage should now be set for the implementation of teacher performance evaluation within the schools themselves.

Implementation of the TPE Cycle

In 1981 Manatt has this to say about the process:

Viewed simplistically, Teacher Performance Evaluation (TPE) is rating, is judging the goodness of teaching. TPE is tough-minded, a quality assurance mechanism, a process performed by principals that compares one teacher to another and to the school organization's standards (p. 3).

In this section we shall look first at the steps in the TPE cycle, with particular emphasis being placed upon Job Improvement Targets. The final section will explore some of the significant factors of TPE that have arisen from research or from field experience.

Steps in the TPE Cycle. The TPE is an integral part of the systems model presented in Figure 4-3. Figure 4-3 shows the flow of activities comprising the cycle that includes both formative and summative types of evaluation. These are the steps in TPE cycle:

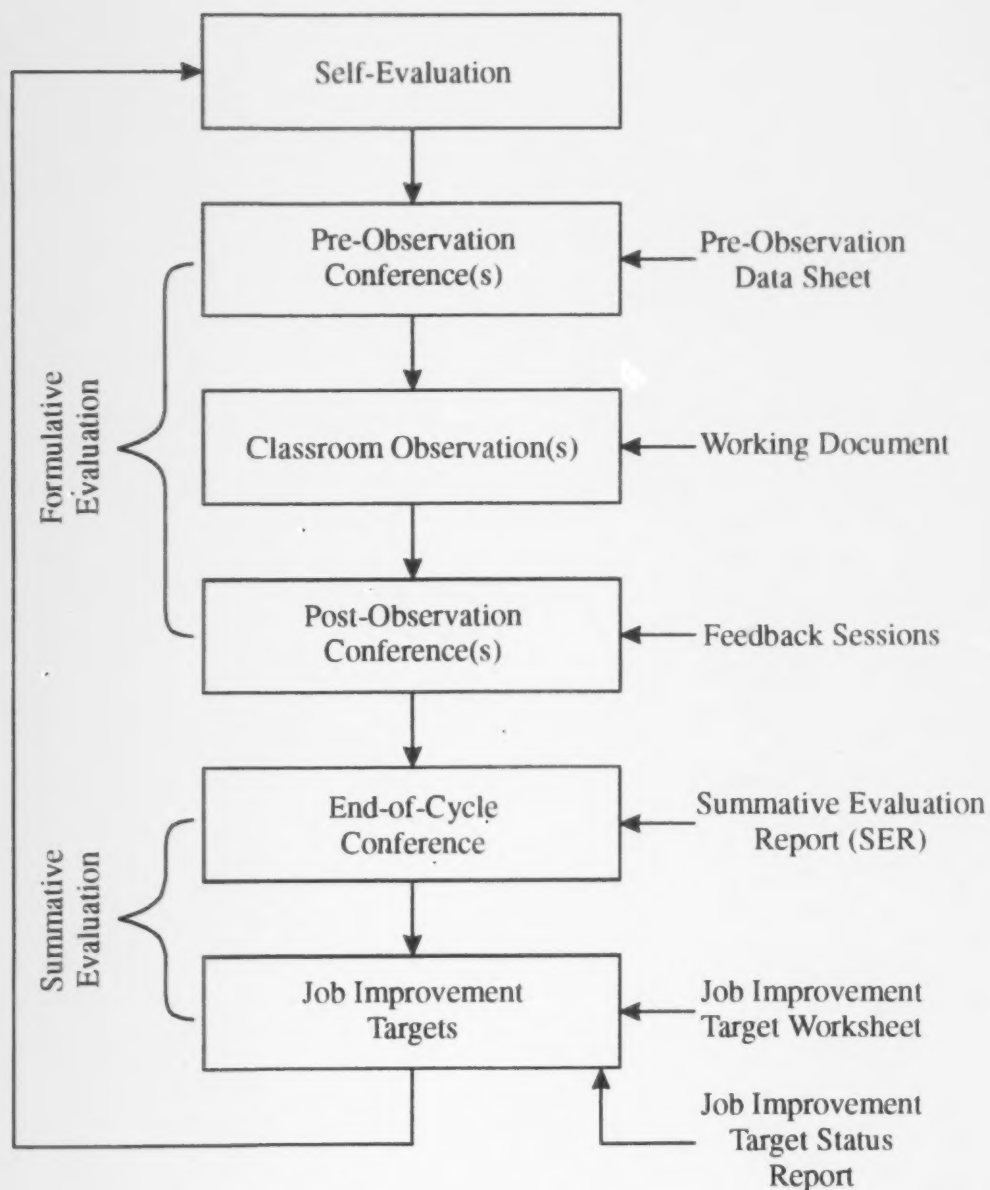
1. establish rules of the game
2. orient teachers
3. analyze lesson plan
4. conduct preobservation conference
5. conduct classroom observation(s)
6. conduct postobservation conference
7. synthesize the data
8. write the evaluation report
9. set job improvement targets

The sequence then repeats itself.

To establish rules of the game and orient teachers. Manatt points out that to be successful TPE requires

- rating scales with criteria based upon effective teaching research
- lesson analysis in conjunction with skilled classroom observation
- coaching and counseling techniques that motivate teachers to change
- provision for procedural and substantive due process regulations to provide protection for both teachers and evaluators

Figure 4-3. Flow Chart of a TPE Cycle



As an essential part of orientation, teachers must be involved with the selection of the performance criteria chosen from a large array of effective behaviors, many of which would have been selected by the steering committee and given the final responsibility for both criteria and standard setting.

Performance criteria will include areas like effective communication with students, demonstrated ability to select appropriate learning content, appropriate management of classroom situations, judicious and effective use of questioning, and so on. Another important aspect of establishing rules and orienting teachers to evaluation revolves around operational procedures such as observations, conferences, and reports. While it is expected that principals will take the lead in these matters, teachers' observations are noted. Neither the teacher nor the principal, however, must be beyond the bounds predetermined by systemwide planning.

A series of descriptors and corresponding response modes is created. Teach mode should contain an established standard level as a guide both to the teacher and evaluator. For instance, classroom management may be described as the maintenance of student interest in an orderly classroom setting; and this may be construed as the standard mode of behavior. Less successful behaviors will include no observable attempts at management, and above standard level behaviors will include the teacher's ability to use a skillful array of approaches to maintain the interest of students at a high level in an exceptionally orderly classroom situation.

Criteria and procedures most often are contained in the handbook used to orient both teachers and evaluators to the process. The important thing is to translate the words of the handbook into vital activities by the personnel concerned. The more closely the rules the game are based on research, the more likely it is that both teachers and evaluator will accept the process. It is here, as much as anywhere, that concerns about the validity of the process are so important. Simply put, the systemwide planning and training periods that were the forerunner to the implementation of the TPE process must have established the validity of the process.

Analyze Lesson Plan. Lesson analysis, also criteria based, is facilitated by the creation of a checklist. Intensive work should have been directed toward the validity of the checklist by the steering committee or more precisely by the performance areas and criteria subcommittee.

Selected criteria for discussion between evaluator and teacher may include those listed below (Manatt, 1981):

- The content, materials, and media selected are appropriate vehicles for teaching the objectives of the lesson.
- The designated instructional procedures are appropriate to accomplish lesson objectives.

- The differences in student capabilities are recognized in the planning of instruction.
- The assessment of student progress on the objectives is indicated.

Conduct Preobservation Conference. Particularly in the case of a teacher new to the system or one starting to undertake new teaching assignments, orientation, lesson analysis, and the preobservation conference are essential to the success of the evaluation cycle. Informal, periodic visits to classrooms will enable the preobservation conference to have meaning. Lunch visits will act as a quality check and commence the focus for formal classroom observations that are to follow.

Before visiting a class the principal will discuss some aspects of the activities that will be briefly viewed. For example, the principal might ask the teacher which particular teaching or learning behaviors are to be monitored or commented on, or which special characteristics of students are to be noted.

The preobservation conference further orients teachers to evaluation, to classroom responsibilities, and to the formal classroom observations that are to follow.

Conduct Classroom Observations. Classroom observations are based primarily upon factors arising from the preobservation conference, although by no means are they limited to these.

There are various ways of approaching classroom observations. The evaluator may use a topical data capture method that will require noting evidence of particular activities. On the other hand, if the clinical supervision approach is used, then the evaluator would need to discern whether the appropriate steps or emphases have been followed. It should not be assumed, however, that all steps have occurred in any one lesson but that they do occur over a series of lessons. Manatt himself (1981) suggested that the following steps should be included:

1. develop anticipatory set, or anticipated lesson outcomes
2. state objectives and why they are important
3. provide input
4. model ideal behavior
5. check for comprehension
6. provide guided practice
7. provide independent practice

Another useful approach to observation is called time line data capture. Each time the teacher changes the concepts taught or methods used, the evaluator jots down the time of day and a brief description of what is occurring. The evaluator elaborates on these comments to the extent that is considered useful for the discussions that are to follow.

The evaluator records observations on two separate forms. The first will contain nonjudgmental and descriptive details useful for immediate advice and improvement. Summative data, recorded on a second form, will be more in line with the extent to which the systems requirements are being met and will correlate with the end-of-cycle summative report form, which was outlined earlier.

Within reason, the more observations that are made, the better. As a classroom observation and immediate feedback are formative evaluation aimed directly at instructional improvement, the usefulness of the feedback increases with the number of observations.

Before the postobservation conference takes place, the evaluator must analyze what has been observed and decide upon the main thrust of the discussion that is to follow.

Conduct Postobservation Conference, and Synthesize and Analyze the

Data. A good starting point for the postobservation conference is to review the decisions made during the preconference. This can then lead to an analysis of the lesson that was observed and questions being addressed such as: What helped learning? Were the key concepts given sufficient time? Was the level of communications satisfactory? Praise should be accorded wherever possible and the whole tone of the conference made positive.

The evaluator will find that some teachers respond positively to straightforward, critical comments, while others will prefer a more indirect approach based upon sharing of perceptions and suggestions. In all probability postobservation conferences will contain an amalgam of both approaches.

At the conclusion of several observations the TPE cycle switches from predominately formative evaluation (nonjudgmental) to summative (judgmental). For some schools this will occur toward the end of the school year, depending upon the district's policy.

The summative report reaches conclusions about how successful the teacher has been during the course of the year; judgments are made against standards prescribed in the handbook. Out-of-classroom performance, as well as classroom teaching form part of the evaluation. The most important reason for the summative report is that job improvement targets, which are generalized teacher improvement goals, are a consequence.

Write the Evaluation Report. As soon as possible, the evaluation report is written. It follows board policy and steering committee decisions. The general framework of the summative evaluation report will determine the format to be followed. To the extent relevant, the principal includes information gathered from informal observation reports. The most important material will come from formal observation reports (with observation notes attached). The report will be based

upon work samples, lesson plans, job improvement targets, posttests and distribution of students' marks, and any other notations that the principal (or other personnel carrying out evaluation) have logged. The keeping of a log or diary of a teacher's activity is essential to make complete and supportable statements in the summative report. In effect, the log is an official business record.

Set Job Improvement Targets. Arising directly from the summative report, a job improvement target commences with a general statement, or goal, of what the principal requires the teacher to do. The job improvement target sets a time limit in which the objective must be attained and contains criteria to be used to measure a teacher's success in reaching the target. Put another way, the job improvement target is a supervisory tool to turn generalized teacher improvement goals into precise and measurable teaching objectives.

Manatt stresses that job targets do not have to be sophisticated but they must be measurable. Job target work sheets are created with space for each of the following components: job targets (including criteria that must be measurable), activities and methods for reaching the objective, comments by the teacher being appraised, and the evaluator's comments.

After the job improvement targets and associated procedures have been set, the cycle recommences. Now classroom observations, conferences, reports, and work samples will be used to *seek evidence of behavior change as well as movement toward the school organization's expectations as outlined in the general teacher performance criteria.*

The job target cycle should lead to a further clarification of roles and responsibilities by both the teacher and the evaluator. Certainly, it should strengthen the commitment of both to reach the targets that have been set. Much will depend upon the increasing ability of the teacher to carry out self-evaluation as part of the process of professional development.

Further Significant Aspects of TPE

Only a brief reference is made here to some of the significant aspects associated with teacher performance evaluation. The intent is to draw the attention of administrators to their importance.

The Principal as Evaluator. Manatt emphasizes the importance of the principal being a role model for teachers—as a counselor, methods expert, clinician, and judge of excellence in teaching and learning. If it is agreed that the process of teacher performance evaluation relies on a measuring of the progress made toward

meeting predetermined performance objectives, the principal, as the evaluator, must play roles in

- classroom observation
- conducting pre- and postobservation conferences with teachers
- analyzing and synthesizing teacher performance and developing advisory and counseling strategies during summative evaluation
- lesson analysis
- setting improvement targets based upon high expectations for continuing evaluation

In all these roles the principal must also ensure that the teacher is oriented to evaluation's potentially positive outcomes.

To play these roles effectively, the principal has to honestly assess and indeed closely analyze various situations of the evaluation process and decide the most appropriate behaviors to associate with each. The task is not easy. For instance, teaching style variables, teacher variables, and context variables have to be considered together with the principal's own biases. Conferencing techniques, legal considerations, methods of handling the marginal teacher, and teachers' expectations of formative evaluation are all difficult aspects of principals' roles as an evaluator and have to be addressed.

In an occasional paper written in 1983, *Competent Evaluators of Teaching: Their Knowledge, Schools, Attitudes*, Manatt provides an excellent account of resources that are available for the training of principals and other administrators as evaluators. The same paper explores which organizational policies promote the evaluation of teaching.

How to Promote Teacher Evaluation. If the concept of teacher evaluation is seen to be important by the school board, the superintendent, and others in authority, its chances of success are considerably greater than the situation where indifference is shown by these people. Sponsorship and promotion are important. So too is the emphasis that is placed on participative planning by involving the kinds of personnel outlined earlier when the School Improvement Model was being discussed. Planning must be complete and must progress and become acceptable by the secure way that it is implemented.

The total context of teacher evaluation must be seen to stand in favor of the best professional practice. If it is known that teacher evaluation aims at raising standards of student achievement, then the association of teacher evaluation for improvement will be recognized as a vital link to student learning.

The Judgment of Teacher Effectiveness Research shows that systematic observation of teachers can produce valuable information for evaluation purposes and also that with adequate training, multiple evaluators can provide information that is more useful than that which the single evaluator can produce. Time and organizational constraints generally result in the principal or a delegated administrator carrying out the evaluation. It should be borne in mind, however, that the TPE approach does allow the involvement of multiple evaluators, particularly when specialized content advice and assessment are required.

The literature on teacher evaluation is replete with discussions on the topic of criteria for effective teaching. The associated problems have been given prominence. However, in recent years major steps forward have been taken in the area of teacher competencies, or teacher performance areas, as well as criteria for judging them. For instance, the School Improvement Model (SIM) Project provides a definitive array of recommended teacher performance areas, criteria response modes, and standards (Blackmer et al., 1981).

In TPE, the end-of-cycle document—the summative evaluation report—should contain, at the maximum, 30 criteria. Most of the items should be focused on classroom management, effective teaching behaviors, and interpersonal relationships. Each criterion should have an allied descriptor of teacher behavior(s) at a required level. It was mentioned earlier that the summative evaluation report form will give ratings on par, below par, or above the standard expectation.

When the final checks have been completed, the tabulated results in report form are forwarded by the principal to the superintendent who is then responsible for summative data being made available to the board of education.

Manatt stresses that TPE should not be used primarily to weed out inadequate teachers, although this may be the outcome of a summative evaluation report, particularly when a teacher has failed to improve after repeated evaluation cycles. The key element of TPE is to help a teacher improve by guided classroom instruction. Even after the complete set of steps for TPE are no longer needed, a teacher may continue to improve by self-evaluation and by having the principal focus the teacher's instructional efforts on a limited number of improvement targets.

Conclusion

Richard Manatt has given new strength and meaning to the practice of Teacher Performance Evaluation. He has achieved this by his very extensive association with school districts and teachers and by conducting workshops and seminars to train principals and other administrators in the various components of TPE. His most important contribution, perhaps, is that he has placed TPE within the complete context of the school district, linking teacher performance to administrator perform-

ance, student achievement, and staff development. The support he has received at Iowa State University has been sustained and nationally beneficial. The School Improvement Model exemplifies these attributes.

Effective teacher evaluation is a vital tool for both teacher improvement and the maintenance of a school district's expectations. As a process it is demanding, never ending, and fraught with potential difficulties. A well-planned TPE approach, where there is thorough commitment by all concerned, helps ensure that teacher evaluation is both acceptable and rewarding. At a time when there is public lack of confidence in education, it becomes a practical imperative that school districts and the community as a whole learn the importance of teacher evaluation and, having done so, use all practical resources possible to strengthen its implementation and processes.

References

- Blackmer, D., et al. (1981). *School improvement model teacher performance criteria with response modes and standards* (Publication No. 81-2). Ames, IA: School Improvement Model Project, Iowa State University.
- Frudden, S., & Manatt, R. P. (1982). *Lesson analysis: The neglected key to teacher performance evaluation* (Publication No. 82-5). Ames, IA: School Improvement Model Project, Iowa State University.
- Manatt, R. P. (1976). *Developing a teacher performance evaluation system as mandated by senate file 205*. Des Moines, IA: Iowa Association of School Boards.
- Manatt, R. P. (1982). *Evaluating teacher performance*. Part I and Part II. [60 minute videotapes]. Alexandria, VA: Association for Supervision and Curriculum Development.
- Manatt, R. P. (1982). *Teacher performance evaluation: Practical application of research* (Occasional Paper 82-1). Ames, IA: School Improvement Model Project, Iowa State University.
- Manatt, R. P. (1983). *Competent evaluators of teaching: Their knowledge, skills, attitude* (Occasional Paper 83-1). Ames, IA: School Improvement Model Project, Iowa State University.
- Manatt, R. P. (1986). *Performance of educational professionals: Lessons from a total-systems approach* (Occasional Paper, 86-1). Ames, IA: School Improvement Model Project, Iowa State University.
- Scott-Walker, R. (1982). *The school improvement model: Tailoring a teacher and administrator performance evaluation system to meet the needs of the school organization* (Occasional Paper 82-3). Ames, IA: School Improvement Model Project, Iowa State University.

Stow, S., & Sweeney, J. (1981, April). Developing a teacher performance evaluation system. *Educational Leadership*, 38(7), pp. 538-541.

Strike, K., & Bull, B. (1981). Fairness and the legal context of teacher evaluation. In J. Millman, (Ed.), *Handbook of teacher evaluation*, pp. 303-341. Beverly Hills, CA: Sage.

Sweeney, J., & Manatt, R. P. (1986). Teacher evaluation. In R. A. Berk (Ed.), *Performance assessment*. Baltimore, MD: John Hopkins University Press.

Toledo School District: Intern and Intervention Programs

During the 1970s, a bitter conflict between the teachers' union and school district authorities in Toledo led to increasing financial problems, teacher strikes, and school shutdowns. Teacher morale, school district credibility, and enrollments altered drastically. What appeared to be a hopeless situation was retrieved by strong cooperative efforts by the teachers' union and school district management resulting in shared decision making in a wide range of educational activities.

One such area was the evaluation of teachers to ensure that quality control was introduced and maintained. In this area the teachers' organization took the lead, thus becoming the arbiter both of definitions of teacher competency and of the professional standards. In this regard, the Toledo system differs from most, if not all other school districts.

The Toledo teacher evaluation model differs in other ways. Importantly, interest is focused mainly on beginning teachers and those whose performance is below required standards. Evaluators are skilled and experienced teachers who receive special training to accomplish their tasks at a high level of competence and acceptability.

It is assumed that once a probationary teacher receives tenure or is not in need of an intervention program, very little attention of the teacher evaluation process is required. In fact, once a teacher qualifies for a continuing contract and appears to be progressing satisfactorily, formal evaluation ceases. The Toledo school district justifies the evaluation of comparatively few teachers on the basis of economies of time, of personnel resources, and financial outlay. In addition, it is considered that teachers most in need of professional improvement are those being evaluated.

Introduction

Perhaps it is only in a city such as Toledo where unions traditionally have been strong that the teachers' union is able to play such a leading part in important aspects

of organizational decision-making, including the evaluation of teachers. It is interesting to note that the one very significant aspect to the resolution of the 1970s conflict of the was that teachers decided that professional development should reside with them. It is difficult to determine whether this was based on disillusionment with previous attempts at evaluation or the knowledge that having control of the evaluation process is a vital component of organizational decision making. Whatever the reason, the approach has received wide acceptance by teachers.

Although the balance of power between the teachers' union and administration remains fragile, a carefully developed scheme has been in operation since 1981. It is one indication of a resurgence of public support for public schools that in 1982 a large levy bond was passed by 70 percent of the voters, the largest margin of support in Toledo's history. Although there is a prevailing view that the shared, cooperative governing of education is working, there is also the realization that it *must work* if present gains are to be maintained. Whatever personal views are held by administrators, there is no doubt that the Toledo Federation of Teachers (TFT) is a powerful authority in decision making. Administrators are able to make few important decisions without at least conferring with the TFT.

One area of strong union influence is the teacher contract. The document specifies not only how teachers are to be employed and what the district's expectations of teachers are, but also states how decisions that affect teachers are to be made. In other words, the contract on the one hand gives stronger than ever protection to the teacher, but on the other makes an unequivocal statement about the teacher's professional responsibilities and required levels of expertise. The teacher evaluation process is a dominant activity in the fulfillment of the spirit of the teacher contract.

The beginning teacher (called an intern) and those who are failing to meet the required levels of competency, and are subject to a program called intervention, are those primarily involved in the Toledo evaluation process. Both the intern and intervention programs, by the nature of their formulation, require close collaboration between unions and administrators. Cooperation is achieved and maintained by a delicate balance of power. There is a realization that the health of the Toledo schools depends upon the strength of this collaborative effort that is central to the decision-making authority of the entire system.

The emerging pattern of teacher evaluation in Toledo is showing a fine balance between a teacher's rights and responsibilities; between top district office administrators' and principals' authority and the union's; and between the school-based and decentralized policy and procedures for the evaluation process.

Responsibilities for Evaluation

There is a centralized structure for collaboration between the teachers' union and school district authorities. The aim of the joint committee is to establish a cooperative continuing evaluation process to improve the quality of instruction. One of its main purposes is to oversee intern and intervention programs (referred to in detail later) by putting into place an Intern Review Board.

All beginning teachers, who are called interns, are evaluated by teachers especially chosen for this task on the basis of their teaching skills, powers of judgment, and experience. The same teachers act as evaluators for any teachers whose level of competency has fallen below required standards. Although the principal plays no hand in the classroom evaluation of first-year interns, he or she will jointly decide with the union's building committee the placement of a teacher to the intervention program. The president of the TFT and the assistant superintendent of personnel must concur with the decision. One reason why the form of evaluation adopted by Toledo has been successful is the wide-ranging support for the scheme. This is shown, in an organizational fashion, by the composition of the Intern Review Board, which is chaired in alternate years by the TFT president and the assistant superintendent of personnel. The principal files an evaluation report on the intern's nonteaching performance at the conclusion of the probationary year.

The principal is given the responsibility of evaluating teachers annually after their probationary year and until they receive tenure. Policy states that the principal should evaluate these teachers once every four years thereafter; but if a teacher gains sufficient qualifications to secure a continuing contract, formal evaluation ceases unless an intervention program is instituted.

The stated aim of the Toledo program of evaluation is to enhance teacher development. Emphasis on counseling for both probationary and intervention program teachers gives a formative dimension to the model. However, it just as clearly serves the purpose of making decisions about a teacher's future. For instance, a teacher will be granted a contract after the probationary internship year only if the evaluation is favorable. Moreover, if a teacher assigned to the intervention program does not receive a satisfactory evaluation, dismissal will follow. The program therefore has a strongly summative dimension, and it follows that *accountability* is an important outcome of the evaluation process. Unlike other educational systems, the Toledo school district has taken the dominant authority for evaluation from the principal and given it to the specialist teacher. Like a growing number of other districts, Toledo realizes the importance of making early and definite decisions about a person's suitability for the teaching profession if incompetency prevails.

The duties of the Intern Review Board, which reports to the superintendent, include policy-making procedures for the continued improvement of the intern and

intervention programs and remediation of any deficiencies that are observed. The Board, which has public recognition, is seen as the chief authority for determining the regulating professional standards. It is interesting to note, however, that the union and management cooperation and decision-making powers may have left the school district board searching for ways to exercise its own authority.

Teachers as Evaluators. By training specialist teachers to be consultants and evaluators and by allowing them to carry out these new duties for a period of up to three years (either full or part time), the school district is making a visible commitment to evaluation. In the early 1980s, when the scheme was commencing, an annual allocation of \$80,000 was given to support the cost of substitute teachers, the training of teachers to be evaluators, and curriculum and other resource materials necessary for the intern and intervention programs. Leaders of the Toledo approach to evaluation maintain that the process is cost efficient, since finances are devoted to teachers needing assistance for their professional development and teacher assessment to maintain the health of the system.

The thoroughness of the activities undertaken by the Intern Review Board is well recognized. As an example, potential evaluators are rigorously screened and trained. The Board also provides ways to assess the quality of the work of these specialist teachers and the credibility of their evaluation reports.

The most important commitment, perhaps, is to the strength of the teaching profession itself. As has been mentioned, the teaching contract defines negotiated work rules that have been developed by teachers themselves. Moreover, the contract document places with practitioners the responsibility of determining who enters and leaves the profession. These are sound reasons why those who are involved in the process are committed to its success.

The principal or delegated supervisor is responsible for first-year teachers who have had some professional experience (and therefore are not called interns), the second year probationary teachers, those teachers who have yet to receive a contract, and all other staff once every four years except those who are qualified for a continuing contract. In practice, many teachers complete 45 months of successful teaching experience, obtain a master's degree, and thereby become exempt from formal evaluation. The possibility of such a teacher falling from grace and having to enter an intervention program is remote.

As the general evaluation program is considerably less important than the intern and intervention programs, it will be commented on only briefly here.

At the completion of the probationary period, evaluation is an infrequent activity. An exception occurs when a teacher returns from inactive status; in this situation evaluation takes place in the same manner as for a beginning probationary teacher.

The General Evaluation Program

Procedures for General Evaluation. The following five steps occur:

1. There is a preliminary conference in which the principal or supervisor outlines evaluation procedures and personal goal setting with a teacher at the beginning of the school year.
2. During the first few months, a teacher is observed and the evaluator assesses the teacher's performance.
3. The first observation conference takes place to establish specific performance goals.
4. There is a deliberate time lapse to allow the teacher to develop these goals.
5. After observations there is a concluding conference in which the evaluator completes the summary evaluation form, which is based on the performance goals that were established as a basis for the evaluation.

First year teachers not involved in the internship program are observed at least 3 times during the cycle and second year teachers at least twice. Each observation period extends for a minimum of 20 minutes. If the initial evaluation results in an unsatisfactory rating, an additional observation period is held. In general terms, attention during the evaluation period is given to teaching procedures, classroom management, knowledge of subject, professional characteristics, and professional conduct. These criteria for judgment apply also to four-year contract teachers. Conforming to school district policy, the contract outlines the procedures for the evaluation.

Principals also evaluate substitute teachers on a regular basis, a procedure that is important in Toledo as these teachers are placed on a priority listing, according to performance, when vacancies occur.

Observations about the Process. Dr. Linda Darling-Hammond, who observed the Toledo approach to evaluation during the early 1980s, commented that teachers thought that evaluation would be improved by

- more frequent observation
- assessment by peers in the subject-matter area, or grade level, of the evaluatee
- emphasis on teaching competence and subject matter knowledge rather than classroom management
- a supportive approach offering guidance in a "continual process of consultation and problem-solving"

These observations are recorded on page 140 of a document that was largely incorporated into a Rand publication, *Teacher Evaluation: A Study of Effective Practices* (1984).

Darling-Hammond also observed that teachers appeared to want a more clinical approach in which a colleague gives advice based upon his or her own area of teaching expertise and comments on classroom problems. Teachers nonetheless seemed somewhat ambivalent about existing procedures. While they saw the necessity of contract provisions allowing protection against harassment, they also sought an improvement-orientated process. As things stand, the general evaluation process is carried out mainly for the purpose of contract decisions in line with stated policy and procedures. To this extent, there has to be a degree of accountability in the process.

Darling-Hammond also observed that principals and other supervisors with a large number of teachers to evaluate neglect the time for proper supervision. As a result, evaluation tends to receive lower priority than the demands of day-by-day management. According to Darling-Hammond: "Standardization of teacher evaluation practices, to the extent that it exists, results largely from due process and grievance procedures," (p. 142). Because constraints are placed on evaluator time and because teachers who perform poorly are rarely improved, the regular teacher evaluation process, protected as it is, is seen by many as a mere formality.

Such is not the case with the intern and intervention programs of evaluation.

The Intern and Intervention Programs

Both the intern and intervention programs are well organized and successful. Highly skilled and experienced teachers supervise and assist both beginning teachers and experienced teachers whose classroom performance is notably deficient. These consulting teachers have the dual task of supervision for improvement and evaluation.

Consulting Teachers. A teacher who becomes both an effective helper of teachers and evaluator of their performance must be outstanding. Apart from teaching skills and requisite experience, they need the appropriate temperament and status in the eyes of other teachers to be considered effective judges. Toledo prides itself in selecting the right teachers as consultants to carry out these duties, which appear to be somewhat in conflict.

During the period of up to 3 years when they are released from classroom duties, the consulting teachers associate with no more than 10 intern or intervention teachers per year. An observation period takes place once every 2 weeks, with more frequent consultations on all major aspects of classroom management, teaching

techniques, and approaches to the teaching of particular subjects. As closely as possible there is a matching of the grade level or the subject teaching level of both the consulting teacher and the teacher being assisted. On those occasions when the subject area match is not possible, the consulting teacher is encouraged to call on the assistance of a senior teacher to help cope with the evaluation of the intern's or intervention teacher's competence in a particular subject area.

Selection. The success of both the intern and the intervention programs depends on the skills of consulting teachers and their professional esteem. The Intern Review Board selects teachers after careful screening. Teachers must have had five years of successful teaching and their applications must be supported by five references, including the teacher's building principal, building representative, and teachers. Selection is based on qualities such as teaching excellence, leadership, classroom management skills, competence in various situations, ability to teach different students in creative ways and, importantly, human relations and communications skills.

An inservice program, initially of three days' duration, is held for consulting teachers. A common framework familiarizes them with appropriate procedures, a range of evaluation techniques, and their role and that of the Intern Review Board.

Role of the Board. The Intern Review Board, which has five teacher and four administrative representatives, assigns consulting teachers, organizes appropriate inservice programs for both consultants and interns, and manages the budget of these programs. Toward the end of the school year the consulting teacher makes recommendations to the Board; these are contained in the teacher's evaluation report. Having considered these recommendations the Board, in turn, sends its recommendations to the superintendent who informs the school board of the decision either to terminate the probationary period of an intern (or the contract of an intervention teacher) or to continue the probationary contract into a second year. Such is the strength of the selection process that the overturning of a consultant teacher's recommendations has been extremely rare.

Implementing the Intern Program. As interns usually have little or no prior teaching experience apart from teaching practicums, some forms of induction into the teaching profession are essential. Too often, however, the importance of assistance to these beginning teachers has been either ignored by school districts or given scant attention. Moreover, when it is considered that only those worthy of entering the teaching profession should be allowed to do so, an organized program of induction, development, and evaluation is essential. Toledo is convinced that it is achieving these worthy ends.

Both the teachers' union and school district management have to agree to the program continuing beyond a particular year. This process allows a review of procedures and a recommitment to policy and practice.

Depending upon the number of interns to be helped, not all consulting teachers' services have been needed each year. Subject specialization also plays a part in assignment by the Board. The Board meets four to six times each year to discuss progress, make adjustments as difficulties arise, give further advice on guidelines, and generally ensure that the program is being implemented as smoothly as possible.

Darling-Hammond observed that consulting teachers have found their work to be "exciting and challenging, and an opportunity for both professional and personal growth" (p. 147). Interns have found the intensive supervision and consequent advice helpful. Their personal skills and self-confidence have grown as a result of the process. The more knowledgeable the consultants in their particular subject area, the more useful and constructive has been the advice given.

Other teachers speak well of the intern program, finding it valuable because it screens out those who would not be good teachers, potentially raises the status of the profession, and eventually should obviate the need for the intervention program.

The question of the authority of a beginning teacher remains a sensitive one in Toledo. The Intern Review Board has ruled that the principals have essentially relinquished control of interns until their second probationary year. While principals are willing to follow the Board's decision, some are seeking further communication with consultants so that they may more adequately be able to evaluate teachers in their second probationary year.

Despite some not unexpected organizational reservations, the goodwill and cooperation extended to the intern program has assured its strong implementation.

Implementing the Intervention Program. Any program designed to improve poor teachers and to dismiss them if predetermined levels of competency are not reached is both bold and courageous. Given the troublesome 1970s, the implementation of the intervention program in Toledo in the 1980s was amazing. The extent to which this implementation has been successful may be attributed directly to the collaboration between union and administrative leaders and the general desire to improve the status of the profession.

A teacher is placed on intervention only if the building committee and the principal agree that it should happen. The identification of the teacher may be made by either the principal or the building committee. To help with this decision making the TFT has published a description of a likely candidate for an intervention program. It stresses that only a teacher who is having *severe* problems with students because of poor teaching skills or inadequate classroom management should be considered an obvious candidate for intervention.

The publication goes on to say that such a teacher will be assigned a consultant to help with the solving of these severe problems. Acceptance of the assigned teacher is mandatory. The document also states that at the conclusion of the intervention period a decision may be made about the future employment of the teacher.

A third year teacher is considered a prime candidate for intervention because, at the conclusion of the third year of probation, a decision is made about termination of contract or the offering of a four year contract. In considering a teacher for intervention reporting, representatives are urged by the TFT to use common sense. Selection decisions must be based on a general recognition that in a particular school a teacher's performance is severely deficient. Obviously, any such decision is very important, since the intervention program may lead to dismissal.

Further precautions are taken about selection. Before discussing an intervention candidate with the principal, building representatives must contact the TFT office. Similarly, the principal must discuss a possible candidate for intervention with the personnel office before taking up the matter with the building committee. When the principal and the building committee have reached a decision that intervention should proceed, formal notification to the teacher comes from the personnel office.

Unlike the intern program, the intervention process has no time limit. Procedures follow much the same format as that for the intern except that the intervention program is far more intensive. Having heard progress reports for an intervention teacher by the consulting teacher, the Intern Review Board decides whether to stop, modify, or continue the program. At any stage the Board may decide to renew or terminate a contract. All observations and discussions are carefully documented, and the consulting teacher may be subjected to intensive questioning by the board as it strives to make the most appropriate decision.

The Toledo intervention program has worked well, although with some initial reticence and even apprehension. The thorough and professional manner in which the process has been conducted has reassured the teachers. The high level of assistance offered to teachers in the intervention program has perhaps been the greatest source of reassurance.

The program has assisted principals in two ways. First, the problem of the poor teacher unwilling or unable to improve his or her performance has been satisfactorily addressed. And second, the difficult tasks of supervising, attempting to improve, evaluating, and possibly recommending dismissal have been removed from the principal's shoulders.

Other Avenues for Teacher Development. Because the intern-intervention evaluation program involves comparatively few teachers, and because other teachers are evaluated rather infrequently or not at all, further avenues for teacher development exist in Toledo. Again, these have resulted from collaboration be-

tween the teachers' union and the school authorities. Included in this category are the School Consultation Program, which commenced in 1978, and the Employee Assistance Program, designed and introduced by administration and union cooperation in 1983. The purpose of both programs is to offer resources to teachers who voluntarily request professional or personal support.

The School Consultation Program provides two teacher consultants, temporarily released from other duties, to give instructional assistance to teachers. Coordinated by a joint committee of board and union appointees, the program has been widely accepted and used by teachers. The purpose of the Employee Assistance Program (EAP) is to offer the services of a full-time counselor to any employee who needs assistance with personal problems that may be affecting his or her professional program. Experience has shown that the EAP in some instances provides either an essential complement, or alternative, to the intervention program. These two programs also support the voluntary staff development programs that offer teachers a wide range and variety of courses for both professional and personal development.

Assessing the Intern and Intervention Programs

As mentioned earlier, a critical analysis was made of the Toledo system for teacher evaluation by Darling-Hammond et al. (1984). One of the most significant findings was that the process extended well beyond the evaluation of minimal competence and indeed went considerably toward attaining the stated aim of establishing by management and teacher cooperation an ongoing process to improve the quality of instruction. The analysis selected three aspects of the process, namely, the validity, the reliability, and the utility to assess the worth of the Toledo intern and intervention programs.

Validity. The validity of the evaluation programs rests on the extent to which they accurately assess teaching competency as defined by the criteria agreed to by union and school district authorities. The central figures are the consulting teachers and their ability to comprehend and assess the teaching practices and standards shown by the teacher being evaluated.

As an essential part of validity, the consulting teacher must judge the accuracy, comprehensiveness, and the appropriateness of the subject content of the lesson as well as the teaching methods used to enhance student learning. Assuming that consensus about content and method exist, the validity of the process is gauged by the extent to which the evaluator assesses the efficacy of these standards during the course of a lesson. The conclusion was reached that the intern and intervention programs generally met the various criteria of validity.

There was also general agreement that consulting teachers have the ability to assess the policy of teaching competency based on prespecified criteria and that, supporting this, consulting teachers are respected by their peers for their expertise in particular teaching areas.

To the extent to which consulting teachers are assigned to evaluate teachers taking lessons outside their specific areas of content expertise, validity falls. This potential fault in the system has been considerably remedied during the past few years.

Another problem that is shared by all evaluation systems is the lack of consensus about what constitutes commonly acceptable standards of practice. To a large degree Toledo appears to have overcome this problem by accepting the teachers' choice of appropriate levels of teacher practices and ways of processing their implementation.

The organization of the Toledo system adds considerably to its credibility. The composition of the Intern Review Board, its policies that work out in practice, and the process for the selection of consulting teachers who have been recommended for a wide range of personal and professional strengths, all imbue confidence.

Reliability. Reliability in the evaluation process depends upon the consistency of methods of measurement across evaluators and across observations. The very nature of teaching makes it difficult to be certain about reliability, since neither teaching nor judging characteristics can be completely anticipated in advance.

Nevertheless, in general terms the Toledo system may be considered reliable mainly because of the strength of the reporting process. To begin with, only a small number of consulting teachers is selected, thus potentially reducing the range of variability. More significant is the regular meeting of consulting teachers with the Intern Review Board, which insists upon a standard and consistent approach to evaluation and to reporting. For instance, a common framework is developed for deciding that the quality of teaching is outstanding, satisfactory, or unsatisfactory.

The frequency of classroom observations strengthens the reliability of evaluation programs. When observations are supported by extensive and detailed consultation, there develops a strong understanding between the evaluator and the teacher about what is being observed and evaluated. Finally, the small number of evaluators in the many schools comprising the system also enhances the reliability as these evaluators develop uniform standards for use in all schools.

Utility. Assessors of the Toledo system viewed its utility from two points of view. First, they looked at how well and how fairly it was measuring what was important to assess; and second, they investigated how well the process was achieving the planned outcomes of the evaluation process without excessive financial and political costs.

An important measure of the program's utility was whether it was succeeding in helping teachers to achieve a predetermined and acceptable level of teaching competency, or warning the system if this level was not being attained. It was found that both goals were being achieved without disruption to the system or harming teacher morale.

The investigators concluded that three critical factors ensured the utility of the intern-intervention evaluation programs:

1. It was seen to be a carefully managed process conducted by evaluators who had no competing responsibilities.
2. It was a tightly focused effort that used limited resources to reach a carefully defined subset of teachers.
3. It was a collaborative effort that engaged the key political actors in the design, implementation, and ongoing redesign of the process.

Continued improvements over the past few years have further strengthened the utility of the process. Its emphasis on consistency of approach, limiting the number of interns being evaluated by a consulting teacher, and focusing on two specific groups of teachers needing special assistance are aspects of the approach that are becoming increasingly acceptable as a cost-effective means of facilitating both teacher improvement and meeting the needs of the organization.

Conclusion

Although it might be argued that the Toledo system fails to reach most teachers and that a program involving so few cannot greatly affect organizational improvement, there is evidence that the intern-intervention program has a favorable spill-over effect. It has changed the character of union and management relationships and enhanced the political climate for the implementation of teacher improvement beyond the evaluation process itself. With the acceptance of the new approach to evaluation and a perceptible strengthening of aspects of the organization as a whole, the collaborative concept of decision making by union and management has been enhanced. Early suspicions and tensions are receding. While disagreements between teachers and administrators exist, the approach to resolving these is changing. The organization, which is led by participative decision making, has tended to overcome problems before they assume too large a dimension.

The approach to evaluation in Toledo may have strengthened teachers' rights, but it also has made them more aware of their professional responsibilities. The intern-intervention programs create and reinforce a professional concept of teaching. Teachers realize that *they* are responsible for the maintenance and development

of professional standards. One likely outcome is that public confidence in the educational system will continue to increase. A tax-paying public requires an acceptable level of teacher competency and mechanisms for school district accountability. Toledo has demonstrated that its approach to evaluation can go a considerable way toward achieving these desired ends.

Note: As an outcome of internal (political) concerns, Toledo School District has very recently discontinued the model for teacher evaluation described here. Nevertheless, the principles contained in the model retain their intrinsic worth.

References

- Coker, H., Medley, D., & Soar, R. S. (1980). How valid are expert opinions about effective teaching? *Phi Delta Kappan*, 62(2), 131-134, 149.
- Darling-Hammond, L., Wise, A. E., & Pease, S. R. (1983, Fall). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research*.
- Fuller, B., et al. (1982). The organizational context of individual efficacy. *Review of Educational Research*, 52(1), 7-30.
- Gage, N. L. (1978). *The scientific basis of the art of teaching*. New York: Teachers College Press.
- Knapp, M. S. (1982). *Towards the study of teacher evaluation as an organizational process: A review of current research and practice*. Menlo Park, CA: SRI International, Educational and Human Services Research Center.
- Millman, J. (Ed.). (1981). *Handbook of teacher evaluation*. Beverly Hills, CA: Sage.
- Rosenshine, B., & Furst, N. (1971). Research on teacher performance criteria. In B. O. Smith, (Ed.). *Research in Teacher Education: A Symposium*. Englewood Cliffs, NJ: Prentice-Hall.
- Shavelson, R. (1973). What is the basic teaching skill? *Journal of Teacher Education*, 14, 144-145.
- Soar, R. S. (1977). An integration of findings from four studies of teacher effectiveness. In G. D. Borich (Ed.), *The appraisal of teaching: Concepts and process*. Reading, MA: Addison-Wesley.
- Wise, A. E., Darling-Hammond, L., McLaughlin, M. W., & Bernstein, H. T. (1984). *Teacher evaluation: A study of effective practices*. Santa Monica, CA: Rand Corporation.

Principal and Peer Evaluation of Teachers for Professional Development

By Anthony Shinkfield

The research-based, positive evaluation methodology for the professional development of teachers, developed by Shinkfield² during the mid-70s, has been applied widely in schools, particularly in Australia. One Australian school using the methodology is St Peter's College, Adelaide, where for the past decade, modified versions of the methodology have been used with notable success.

Although one attractive feature of this approach is its flexibility, which has given rise to an imaginative array of procedures according to a particular school's context, the basic principles underlying the approach do not vary. An implicit assumption of the approach is that improved teaching performance will be beneficial for student learning. The guiding principles are as follows:

1. There must be acceptance of teacher evaluation within the school as an integral part of educational process.
2. In the process of evaluation, teacher development will occur only if a constructive approach is followed.
3. Collaboration and mutual respect between teacher and evaluator are essential.
4. General agreement among concerned parties about the school mission and job assignments must precede implementation of a plan of teacher evaluation.
5. Teacher self-appraisal must become a significant part of the process.

Personnel evaluation has always been important in order to meet demands for teacher accountability. It is hoped, however, that educators will give increased attention to using evaluation of teacher performance to improve instruction. The St. Peter's College method holds firmly to the fundamental principle that personnel evaluation is not only for accountability but must also be an integral part of the educational process within any school. While formative evaluation for professional development may, in certain situations, be replaced by summative evaluation for accountability, the initial emphasis should reside in the former role of teacher evaluation.

² A South Australian with extensive experience in educational evaluation, Anthony Shinkfield is a member of CREATE'S National Advisory Panel.

The theory underlying the model described here stresses *positive* appraisal techniques that include the emphasis being given to self-evaluation, constructive feedback, an open climate during discussions between evaluator and teacher, and a strong commitment by both to a teacher's professional development.

Although the approach has been used mainly in independent schools like St. Peter's College, there is no reason why it should not be appropriate to public schools. In either situation, evaluation must be grounded in advance agreements and collaboration between the concerned parties. A handbook of procedures must demonstrate the involvement and support of individual school councils, teacher associations, teachers themselves, and principals.

Responsibility of the Principal as Evaluator

The St. Peter's approach assumes that the school principal will be the evaluator, or one of two evaluators, and also chairman of the "Assessment Committee." The composition of this committee is addressed in the next section. It may be said, with justification, that many principals are ill-equipped to assume these tasks without the help of specific training. The importance of evaluator credibility is also referred to later.

The stance taken here is that the evaluation of staff is one of the most essential responsibilities of any school principal. Whether trained or not in the various areas of instructional responsibility, the school principal must examine the performance of staff members in order to provide constructive feedback and to make decisions that affect individual teachers and the school itself. While some delegation of administrative responsibilities is appropriate, the ultimate responsibility for the professional development of staff, including evaluation and curriculum matters, must reside with the principal.

Even if principals are not knowledgeable in evaluation techniques, they can quickly develop a sufficient level of proficiency in evaluation if they give it high priority. Study of judiciously selected evaluation writings is useful but perhaps not as important as a strong personal motivation by a principal to reap the organizational benefits of an ongoing, positive personnel evaluation process that is accepted by staff.

The Assessment Committee (or Panel). The St. Peter's method essentially involves an Assessment Committee consisting of three persons: the teacher being evaluated, a peer nominated by that teacher, and the principal.

As the evaluation is designed to extend throughout a school year, and as the principal's time is limited, it has been found that a school may not be able to evaluate more than four teachers each year unless other administrators are involved. For this

reason, from time to time other administrators (for example, a deputy principal) may replace the headmaster/principal as the chairperson of the Assessment Committee.

With such administrative support, it should be possible to evaluate most members of even a larger school staff once every two years. Much depends, however, on the commitment of those who are involved, including their willingness to undertake some training and evaluation, and the total cohort of evaluators that a school is able to develop.

In practice, it is more usual to concentrate initially on beginning teachers and those who are more experienced teachers who seek or need redirection or remotivation for teaching. At St. Peter's College, a beginning teacher's contract states that evaluation will occur for induction and professional development purposes. The contract also states that a review after one year will determine whether tenure will be offered. The emphasis, however, is on the positive nature of evaluation as defined earlier. Experienced teachers, particularly those whose instructional methods and curriculum knowledge have changed little with the passage of years, may need more persuasion to undergo evaluation. Nonetheless, this group too has shown increasing willingness to participate in teacher evaluation, particularly after the process was seen to be advantageous for their peers.

For the sake of simplicity, it will be assumed that the usual situation prevails in which the principal is the chairperson of the Assessment Committee. The word "peer" encompasses any third staff member chosen by the teacher being evaluated. In the St. Peter's situation, the peer has been a fellow classroom teacher, a head of subject department, or a grade-level supervisor.

Evaluator Roles, Training, and Credibility. Evaluator roles, training, and credibility are three interconnected functions of the evaluation process. Those participating in evaluation must know the role they are to play including, importantly, their responsibilities.

In the model being presented, policies play a vital part in role definition. Just as the design of the system must be both clear and specific, so too must be the roles of the principal (or whoever is chairperson of the Committee), the teacher being evaluated, and the third member of the Committee. It is important, in the St. Peter's situation, that the teacher being evaluated completely understands that self-appraisal and constructive outcomes of evaluation depend upon the teacher himself or herself. While the whole process will help an individual teacher to improve professionally, motivation must be an essential part of the role played by the teacher.

The principal, as chairperson of the committee, must obviously be thoroughly conversant with policy and procedures and also be able to explain these to a teacher in a convincing fashion. It is essential also that the third member of the Committee understands and supports the evaluation process and its intended outcomes. Pre-

dominant, however, is the leadership role of the principal, since this person must ensure that an Assessment Committee forms and functions in a satisfactory and positive fashion.

It has been found that a *minimum* training period for principals, or others who are to lead Assessment Committees, is a complete day of inservice work. During this time basic principles of evaluation are presented, the Shinkfield model offered as one approach if a formative evaluation emphasis is required, and sessions offered in such aspects of evaluation as conferencing, observational skills, analyzing and synthesizing information from observations and other sources, and report writing. Relevant evaluation materials are sent to conferees approximately a week in advance of the inservice day. The value of reading and attempting to understand these materials is always advised. There is no doubt that the potentially good evaluators are those who give high priority to the importance of training as an adjunct to the implementation and development of teacher evaluation in their own schools.

Any evaluator's credibility is increased by the strength of assessment policies, the cogency of the design or model that is to fulfill aspects of the policy, and the clarity and specificity of the procedures that are used. Principals who have adopted the St. Peter's model for teacher evaluation have been strongly urged to set a credible foundation before any evaluation process commences. If they know precisely why teachers are being evaluated; whether procedures are appropriate, justified, and accurate; and whether instruments for gathering information will provide consistent indicators of the performance of teachers being assessed—to name some of the many important aspects of teacher evaluation—then strong starting points have been established.

With these assumptions in mind, we move on to the various stages, principles, and procedures of the approach. This section will include evaluation strategies designed to promote individual professional growth. It will take into account different professional development needs, career stages, teaching context and the creation of motivation needed for change to occur.

Stage One: Climate and Policy

The essential foundation for the success of any teacher evaluation program is that it must be seen in a positive light by all those involved. Firmly establishing an improvement orientation, then, is of vital importance. This responsibility resides with the principal who must communicate to staff the importance of a successful teacher evaluation program for the continuing health of the school and for promoting the success and positive self-image of each professional in the school. At St. Peter's this has been achieved basically through a number of staff meetings and

subsequent inservice courses with small groups participating. Although this task may appear daunting initially, once staff are involved and can see the purpose of what is being promoted, acceptance becomes a real possibility.

If the principal is not personally convinced about the importance of evaluation or is hesitant or unenthusiastic, staff will quickly become either disillusioned, skeptical, or disinterested.

The principal must emphasize

- that personnel evaluation is designed to help both the individual and the institution
- that a professional, open climate will prevail during all stages of the evaluation process
- that teacher job satisfaction will be heightened, and student learning skills increased
- that the prime purpose of the evaluation is for teacher improvement and not for teacher's future employment prospects
- that, in keeping with the professional nature of the evaluation, collected information and discussions will be kept confidential

If teachers realize that they are accountable both to their students and their organization, they should welcome a thorough review of their progress. If this is coupled with a general staff feeling that an evaluation is worthwhile, satisfactory outcomes are assured.

Written Evaluation Procedures. Written documentation is essential in the St. Peter's situation. Almost a decade ago, representatives of the staff association and administration met to agree upon policy and procedures. It was agreed by all concerned that these would be straightforward and as brief as possible. It was also decided that they should be flexible to make sensible changes. In fact, a number of modifications have been made over the years, but the general precepts of the guiding model have not altered. Several schools, following a similar approach as St. Peter's, have included teacher evaluation procedures in staff handbooks.

In general terms the St. Peter's staff evaluation documentation follows the stages outlined in this chapter. For this reason and more importantly because such documentation must be significantly and closely allied to the context of a particular school and its needs, a sample of the St. Peter's documentation is not given. In any case, as has been mentioned, some changes, particularly in forms and documentation, have been made to meet new circumstances or to acknowledge that the school's grasp of personnel evaluation has strengthened with practice.

It is interesting to note that over the years, with the success of the approach and accumulated knowledge, credibility for teacher evaluation has also increased.

Although it is difficult to prove, there is a feeling that teacher evaluation has positively influenced other aspects of school life beyond the teacher's own classroom.

Once policy documentation was completed at St. Peter's College, initial conferences were held with teachers who volunteered to be the first to undertake the evaluation process. In most subsequent years, there have been more teachers requesting evaluation than could be accommodated.

Stage Two: Initial Conference(s)

It is probably true that the initial conference can make or break the potential success of the evaluation process. Very few of us wish to be subjected to critical appraisal, and most become defensive, particularly when imperfections begin to be exposed to the light of day. It is insufficient to simply emphasize that the confidentiality, at all stages, will be assured. The principal must develop a feeling of mutual respect between teacher and evaluator and a sense that an exciting professional enterprise is being undertaken for the benefit of the teacher and the school. To achieve these ends, the principal must carefully consider the strategy and words to be used, bearing in mind that an approach applicable to one teacher may be inappropriate for another.

The First Meeting. As has been mentioned, the teacher is given the right (which invariably has been accepted) to choose the third person who, with the teacher and the principal (or a deputized administrator) comprise the Assessment Committee. The offer is made at the first meeting with the principal and, if accepted, the third person will be present at most subsequent conferences. It has been found in practice that the teacher sometimes appreciates the advice of the principal about who best can provide support and professional expertise for effective evaluation to proceed. Nonetheless, the final decision is made by the teacher.

The chairperson of the Assessment Committee must make abundantly clear the purpose for the evaluation. The aim is teacher improvement. Under no circumstances must the chairperson allow his or her biases to predominate or even appear during this first meeting. If the teacher knows that the purpose of the evaluation is professional improvement, then there need be no endeavor to hide potential weaknesses nor in any way to "whitewash" the outcomes of classroom observations.

Next, the principal outlines procedures that will be adopted during the remainder of the evaluation. These are based directly on the documentation for teacher evaluation procedures that exist at a particular time. It is important that the principal set aside sufficient uninterrupted time for all of the teacher's concerns to be

addressed in a thorough fashion. At the end of the conference, the teacher must be convinced that the process will be advantageous both personally and institutionally.

Second Meeting. The second meeting, at which the complete panel is present, closely follows the first. This meeting has two main purposes:

1. to emphasize the importance of self evaluation
2. to lay the groundwork for one of the most significant aspects of the evaluation process, namely, written statements about a teacher's strengths and weaknesses

These two evaluation elements are closely linked, often inextricably entwined.

The principal as an evaluator, or indeed others involved in the teacher assessment process, fundamentally act only as catalysts in the important matter of teacher improvement. It is the teacher who is responsible for his or her own professional development. For this reason, self-appraisal is a vital component of all that follows. Where teachers see the importance of wishing to control their own professional growth, evaluation is seen as an integral part of the process.

The teacher is asked to consider teaching strengths and weaknesses; the teacher's claimed strengths and weaknesses then form the basis of procedures from that time. Moreover, the principal states that the other two members of the Assessment Committee will also formulate lists of the teacher's strengths and weaknesses. The principal should comment that those lists compiled by the chairperson of the committee and the peer person will most likely be more general and less objective than those that are thoughtfully made by the teacher. Practice has shown that when the three lists of strengths and weaknesses are compared, there is considerable overlap. If nothing else, a basis is formed for a substantial list of weaknesses, about which performance objectives may be drawn, and strengths that give the teacher the assurance that much has already been achieved.

A couple of things occur at this second early meeting. The St. Peter's College Teacher Competency and Duties List is handed to the teacher as a guide in the formulation of strengths and weaknesses. This Competency and Duties List, which is displayed in Table 4-1, is an adapted version of that produced by Redfern in 1980 (pp. 21-23). The teacher is told that this Competency and Duties List will form the major part of the next stage in the evaluation process. The teacher is also requested to align perceived strengths and weaknesses with the terms contained in the duty statement given to the teacher at the time of employment. Each teacher's duty statement contains the school's expectations with respect to classroom procedures, involvement in school activities beyond the classroom, and an understanding of the school's traditions and culture.

Table 4-1. St. Peter's College Teacher Competency and Duties List—A Teacher's Responsibility Objectives Stated as Criteria

1. The School Culture

- 1.1 Has knowledge of the basic aims of the School as stated in Standard Procedures.
- 1.2 Uses a knowledge of the School's history and traditions in classroom situations.
- 1.3 Makes appropriate reference to the School Rules to remind students of their responsibilities and the necessity for an orderly way of life among all members of the School Community.
- 1.4 Follows set procedures for morning classroom prayers and other rituals as appropriate to a School with a Church of England (Anglican) foundation.
- 1.5 Understands and makes use of a student referral system.

2. Planning and Organizing

- 2.1 Makes short- and long-range curriculum (syllabus) plans.
- 2.2 Correlates individual objectives laid down by the Curriculum Committee and individual Departmental Committees.
- 2.3 Adheres to the principles laid down by the Curriculum Committee and individual Departmental Committees.
- 2.4 Plans appropriate sequence of skills.
- 2.5 Has an ongoing program to diagnose and assess the needs and progress of individual students.
- 2.6 Adjusts physical environment to accommodate variety in learning situations as may be appropriate from time to time.
- 2.7 Co-operates with others in planning daily schedules of activities both within and without the classroom.
- 2.8 Manages time efficiently.
- 2.9 Keeps accurate records.
- 2.10 Adheres to all procedures laid down in Standard Procedures for planning and organizing.
- 2.11 Prepares reports that reflect accurately the progress of students.

3. Motivating Students to Learn

- 3.1 Motivates by positive feedback and praise.
- 3.2 Is responsive to the needs, aptitudes, talents, and learning styles of students.
- 3.3 Develops learning activities that are challenging to students.
- 3.4 Provides opportunities for student expression in a variety of ways, both spoken and written.
- 3.5 Stimulates students to participate in class discussions and activities (emphasis here is on participation by *all* students).
- 3.6 Generates a sense of enthusiasm among students.
- 3.7 Helps students experience social and intellectual satisfactions.
- 3.8 Relates curriculum to situations both within and without the School.

Table 4-1. (continued)

3.9 Stimulates participation by judicious use of questioning.

4. Relationships With Students

- 4.1 Stimulates an orderly, disciplined student atmosphere by:
 - 4.1.1 anticipating instances of undiscipline
 - 4.1.2 acting immediately to restore sensible discipline
 - 4.1.3 maintaining a friendly classroom climate (never speaking down to students)
 - 4.1.4 being fair and consistent in methods of punishing acts of undiscipline
 - 4.1.5 showing warmth and understanding in these dealings.
- 4.2 Collects pertinent information about students and maintains the confidentiality of this documentation.
- 4.3 Counsels students individually and in groups.
- 4.4 Promotes an open but controlled atmosphere, enabling students to express their opinions.
- 4.5 Helps students develop positive self-concepts.
- 4.6 Encourages students to define realistic goals for themselves.
- 4.7 Shows concern for students who have personal problems or handicaps.
- 4.8 Encourages students to strive for high achievement and excellence according to their abilities.
- 4.9 Utilizes the resources of student personnel staff services (Housemasters, Chaplains, Career Master) as appropriate.
- 4.10 Makes self available for conferences with students (for personal or collective problems, or concerns, or clarifications).
- 4.11 Guides students in the observance of fair and democratic principles.
- 4.12 Manages behavioral problems on an individual basis and attempts to solve these personally while recognizing the necessity to seek help from administrators where appropriate.
- 4.13 Has a sound rapport with students while maintaining a professional social distance.

5. Utilizing Resources

- 5.1 Is aware of available resources in the Library and other places as appropriate.
- 5.2 Uses a variety of available resources.
- 5.3 Uses the physical environment of the School (both buildings and grounds) to support learning activities.
- 5.4 Adapts available resources to individual needs of students.
- 5.5 Uses the services of specialists (e.g., Heads of Departments) in the selection and utilization of resources.
- 5.6 Uses equipment and materials effectively.

6. Instructional Techniques

- 6.1 Encourages students to think.

Table 4-1. (continued)

-
- 6.2 Uses a variety of teaching techniques.
 - 6.3 Uses a variety of instructional materials.
 - 6.4 Varies opportunity for creative expression.
 - 6.5 Helps students apply their experience to life situations.
 - 6.6 Conducts stimulating class discussions which are always under the control of the teacher and firmly directed by that person.
 - 6.7 Encourages the development of individual interests and creative activities.
 - 6.8 Uses appropriate assessment techniques to measure student progress.
 - 6.9 Assists students to evaluate their own development in a particular subject.
 - 6.10 Enables students to share in carrying out classroom activities.
 - 6.11 Shows flexibility to carrying out teaching activities.
 - 6.12 Creates an atmosphere of mutual respect between students and teacher.
 - 6.13 Enables students to learn how to work independently and in groups.
 - 6.14 Promotes group cohesiveness and team spirit.
 - 6.15 Uses questioning to involve all members of the class and to evaluate extent of students' grasp of old and new concepts.
 - 6.16 Develops an interesting array of techniques (both formal and informal) for the assessment of student work.

7. Relationships with Parents

- 7.1 Makes adequate use of parent evenings.
- 7.2 Encourages parents to discuss problems relevant to their children.
- 7.3 Interprets learning programs to parents.
- 7.4 Stresses a positive approach in School/parent relationships.

8. Professional Development and Responsibility

- 8.1 Participates in the development and implementation of School policies and procedures.
 - 8.2 Maintains good rapport with colleagues.
 - 8.3 Keeps self up-to-date in areas of specialization.
 - 8.4 Takes advantage of inservice education opportunities.
 - 8.5 Participates in School and systemwide professional activities as they become available.
 - 8.6 Assists in out of class activities (games, clubs, etc.).
 - 8.7 Shares ideas, materials and methods with professional colleagues.
 - 8.8 Consults with students' previous teachers, Heads of Departments, and visiting consultants to improve the teaching/learning process.
 - 8.9 Interprets School programs to parents and to the School Community as opportunities occur.
-

Stage Three: Competency Objectives

After approximately two weeks, the teacher to be evaluated meets again with the other two members of the evaluation panel. A comparison is made between lists of strengths and weaknesses that have been drawn up and, following a thorough discussion, a final list of strengths and weaknesses is selected. As the final list will be based primarily on the teacher's own considered preferences, deference must be given to the list compiled by that person. In general terms, it has been found that neither final list will contain more than ten items and very often fewer than this. Such a list should not contain a detailed account of minor teaching faults that the teacher may perceive to have. It is quite possible that these will be remedied during the course of the evaluation.

At this stage, reference is made once more to the Teacher Competency and Duties List, which has been furnished to the teacher being evaluated and the third member of the evaluation panel.

Teacher Competencies and Duties. Irrespective of a formal evaluation process, such as the one under discussion, a school staff should produce a list of teacher competencies and duties that underline sound student learning and development. If such a list has been compiled at staff inservice, it is an extremely valuable adjunct for teacher evaluation. In the St. Peter's College context, it is essential.

Table 4-1, the St. Peter's College Teacher Competency and Duties List, indicates that a wide range of teacher competency areas is considered during teacher evaluations and on other occasions, together with duties that all teachers are expected to fulfill. This list is never complete nor fixed. Changed circumstances and review of the list at subsequent staff professional sessions will see it modified and further aligned to student needs and school community expectations.

The teacher competencies and duties contained in Table 1 are actually a list of teacher activities developed under general concepts. For example, it can be seen that the concept "Learning Strategies" encompasses such detailed competencies and duties as diagnosing student needs, selecting appropriate resources, designing appropriate instruction, and evaluating effectiveness, among other relevant activities.

The panel must now spend considerable time defining, as closely as possible, the various weak (or relatively weak) areas finally selected. These, in all probability, will coincide with some of those contained in the school's teacher competency and duties list, except that they will be expressed negatively. The next and more difficult task is to provide a brief, written example of each competency. At St. Peter's College, the principal has found it advantageous to have some of these prepared in advance, even though it must never be assumed that the principal's selection of areas needing strengthening will dominate that of the teacher being evaluated. The

important thing at this conference is that the teacher is satisfied with the selection of areas to be improved and that he or she knows how to go about it. The teacher has also been made aware that other weaknesses may be discerned by evaluators during the classroom observations that follow. Such areas will then need to be defined as teacher competencies and duties and written examples of each provided as criteria for advice and improvement. In other words, the teacher must be certain about what actually constitutes suitable levels of teaching attainment or duties fulfillment in particular areas. Thereafter, the defined teacher competency or duty, supported by a description of satisfactory level of attainment, becomes a Standard for teacher evaluation. The development of these criteria, as standards for judgment, also may be undertaken as part of staff professional inservice.

Different competencies and duties will be used or stressed according to different kinds and levels of preparation and experience of individual teachers. Let it be emphasized again that teacher collaboration should be sought when decisions are being made about which competencies and duties are more relevant and which should ultimately be used as the basis for evaluation. It has been found, for example, that a teacher with some experience may not need to have many of his or her classroom competencies assessed, but may need to have knowledge about a duty, such as curriculum change in a particular discipline area, evaluated closely.

At the completion of this conference, the teacher is furnished with a written set of expectations for improvement, a copy of which is placed in the teacher's evaluation file. Other copies are retained by the principal and the third member of the Committee. Steps are taken to ensure that all copies are held confidential.

Classroom observations, professional judgments by those evaluating, immediate feedback, and consequent advice and decisions about improvement are the main components of the model, which is basically formative in nature. As pointed out earlier in the chapter, in certain situations where minimum performance standards are not met or dereliction of duty occurs, summative evaluation for accountability will occur. However, in a situation like St. Peter's College, where rigorous selection procedures apply, formative evaluation predominates almost exclusively. Such a situation has been termed "elitist."

Formative Evaluation. Formative evaluation consists of selecting appropriate information for systematic and continued revision. It is important to show teachers how to change or develop. Feedback to the evaluated teacher, the essence of formative evaluation, has immense potential worth for teacher development and improvement. The teacher is not an indifferent bystander in this process. He or she must be deeply involved in discussions, being constantly aware that self-appraisal, self-help, and improved skills are essential parts of the evaluation process and that the improved quality of education in the school is a product.

Stage Four: Observations

The teacher selects the lessons to be observed in the first instance. Thereafter, the teacher is observed in different classroom situations. Mutual agreement is reached between the teacher and evaluator regarding the time of classroom visits. These are carried out by the two evaluators. It has been found useful in the St. Peter's context for observations to occur approximately once a month and to continue, if necessary, throughout the school year. Classroom observations constitute the main thrust of the evaluation system and therefore must be scheduled on a regular basis.

The principal and the third member of the evaluation panel may have different roles to play in the observation process, as well as overlapping areas of responsibility. For instance, if the peer person is a subject specialist, observations, judgments, and decisions about advice for the teacher may be based on the content of what is being taught. On the other hand, the principal may be focusing on areas like communication between teacher and student, aspects of questioning, skills in classroom management, and the like. One evaluator complements the other, and both use their observational skills to help the teacher improve.

During each session, the evaluator sits in an unobtrusive position in the classroom (usually at the back). It is remarkable that students quickly become used to the presence of an "outsider," particularly after the first visit. Because the situation is unnatural, the evaluator must do everything possible to put the teacher at ease and to reduce any anxiety. A great deal of the goodwill that has been deliberately developed to this point will quickly dissipate if the teacher finds the situation at all threatening. It has been found advisable to keep note taking to an absolute minimum while maintaining a maximum concentration on observing what the teacher is actually doing and saying.

The evaluator then makes time available immediately after the lesson to record observations. A method that has been found effective is to make notes based on observations under each of the competency objectives that have been the focus of the classroom visit.

A follow-up conference between the evaluator and the teacher takes place after the conclusion of school on the same day as the observation.

Follow-Up Interview. The commencement of a follow-up conference will be a sound gauge as to whether the process is working well. The evaluator will have in mind such questions as: Is the teacher at ease? Will there be defensiveness? Have I personally observed well or are there significant factors that have been overlooked? Does the teacher sincerely wish to improve as a result of the evaluation procedures?

To obviate what may be potential concerns, the evaluator begins on a positive note by complimenting the teacher on some particular aspect of what has been

observed and in other ways indicates that there is a desire to give professional support to the teacher.

Any areas of concern should be approached by judicious and sympathetic questions from the evaluator. If information is sought in this way, it has been found that an open climate is established and a foundation set for profitable discussions and decisions. It is important to minimize criticism, as this threatens the teacher's self-esteem.

The evaluator takes notes immediately after this postobservation session. These will form the basis of the second preobservation conference.

Second Postobservation Conference. The second postobservation conference is a crucial factor in the success of the St. Peter's College model. Present are the teacher, the principal, and the third member of the evaluation panel. As an outcome of the two postobservation conferences that have been held (that is, between the teacher and the principal, and the teacher and the other Committee member) the following occurs:

1. Any further strengths discerned are added to the appropriate list.
2. Where performances have been satisfactorily met, these are deleted from the weakness list.
3. If appropriate, further competency objectives are developed in the same way as earlier.

Again, the teacher is furnished with a brief, written account of what is to transpire as an outcome of the formative evaluation process. During this conference, the school's Teacher Competency and Duties List may be referred to once more, as well as the teacher's job description.

According to the progress being made by the teacher, it may not be necessary to have a realignment of goals or objectives after each series of classroom observations. Sometimes this aspect of the evaluation cycle takes place after every second series of observations. Whatever occurs, the teacher must be informed explicitly, be clear in his or her mind what is to occur, and be caught up in the professional excitement of knowing that improvement is actually occurring.

Stage Five: The Wind-Up Conference

The wind-up conference completes the planned cycle of teacher evaluation. Depending on the progress made by the teacher, it may or may not conclude the need for formal evaluation. Certainly, it will not be the end of the evaluation from the point of view of the teacher's own self-appraisal of teaching performance, success

in developing student learning, and strengthening worthwhile aspects of the school itself.

Decisions to be Made. Prior to the wind-up conference, the two evaluators confer and arrive at the final evaluation report. This is based upon all teacher competency and duties objectives having been addressed, including those that have been satisfactorily accomplished; those that have been modified; and any new ones that have been added during the course of the evaluation. It is signed by both evaluators and marked confidential. The teacher is given a copy of this before the conference.

When the session commences, the principal requests the teacher to respond. It has been found in practice that the teacher's responses have invariably been positive, since the final report contains few, if any, surprises, and because problems have been addressed as they have arisen during the cycle of the evaluation. This does not say, however, that problems may not still exist. If they do, then it is possible for the evaluation to continue at an appropriate time. In the St. Peter's College context, this happens in the following school year, provided, of course, the teacher wishes to continue in the school and the teaching profession.

Although the situation has not occurred at St. Peter's College, another school using the same approach to evaluation has found that a beginning teacher reached the conclusion that for him personally the difficulties he encountered in teaching were too hard to overcome. It is worth noting that the open climate in which the evaluation has been conducted and the evident professionalism inherent in the procedures and by the personnel involved ensured that the teacher left with positive feelings toward the school and the teaching profession itself. The teacher's decision to leave the profession was not anticipated when the evaluation took place; the fact that the teacher left without impairment to his self-esteem says much for the way in which the evaluation was conducted.

Continuing Self-Evaluation. The wind-up report should begin with positive aspects of the teacher's professional skills. It also should suggest ways in which improvement may take place. It has been found most useful during the wind-up conference for the school principal to reiterate that any durable and continuing professional improvement must reside with the teacher.

Apart from advice given by the principal and the third member of the evaluation panel, the teacher will receive considerable help from the school's own professional development library. The principal should be sufficiently aware of the contents of this library to be able to direct a teacher to appropriate journal articles or books.

If, as a result of the evaluation, the teacher not only has increased awareness, confidence, and skills in areas of previous concerns, but also has a heightened belief in the correlation between these personal improvements and benefits to student learning, then the evaluation has progressed very well. It is an added bonus if both

teacher and evaluator are able to conclude that the school also has benefited as an organization. In the case of beginning teachers, it may take several cycles of evaluation for this conclusion to be reached.

A Miscellany of Important Aspects of the Approach

A number of factors are mentioned briefly in this section. Over the years, they have been important to the success of the approach outlined.

Leadership and Motivation. Most teachers have high self-esteem and a desire to participate in their own performance assessment. It follows that the higher the expectation a principal has of a teacher's ability and worth, the more chance there is of the teacher evaluation process succeeding. There must be an acknowledged confidence and trust between the principal and teacher and indeed between all three members of the evaluation panel. The principal is wise to defer to the teacher, wherever possible, to solicit ideas and opinions. Sound communication practices follow from active encouragement from the principal and the other panel member for the teacher to participate in an active fashion in all aspects of the evaluation.

The principal must realize that, as the school leader, he or she has the ultimate commitment to develop the human resources available in the school organization. None is more important than the teacher. If the principal is able to so set the conference climate that there is a stimulating dialogue regarding strengths, weaknesses, and improvement plans, then there are strong chances that desired outcomes will be attained.

Teacher Responsibility. Mention has been made several times in this chapter of the importance of the teacher assuming the responsibility for his or her own professional development. It is sometimes necessary to emphasize to a teacher that improvement will not simply happen or that a consistent, conscientious, and occasionally painstaking effort is necessary before desired ends are reached. When the evaluation cycle has been completed, the teacher will have received considerable support from the principal and school. The groundwork has therefore been laid for continuing self-appraisal and self-improvement with the basic initiative coming from the teacher, supported, of course, by others in the school as well as professional literature.

There are rare occasions where a teacher may not respond to the very best efforts of those directing and helping with the evaluation. If a subsequent evaluation cycle is equally unsatisfactory and thorough documentation has indicated that attempts have been made to help the teacher, it is the principal's task to advise a change in vocation. Constructive criticism, accepted and followed professionally, will usually

overcome such a situation satisfactorily. If this does not occur, formal summative evaluation is needed to assess whether the teacher has attained minimum standards required by the Competency and Duties List (stated as criteria).

Evaluation for Retention. The evaluation approach promoted in this method is formative. This carries the expectation that a teacher will wish to improve performance and, indeed, is proud of his or her profession and school. Thus, the approach fortifies a positive concept of evaluation.

A different evaluation role is necessary if the principal needs to assess a teacher for retention. This is summative evaluation. The significant difference from a formative evaluation is that the principal or school now establishes the objectives against which the teacher will be evaluated.

There is no reason why a school should not incorporate both formative and summative evaluation processes among its activities and have these documented in the teacher handbook. There is *every* reason why one method should not be mistaken or confused with the other. It is essential that the differences in procedures and intentions are thoroughly clarified even though both formative and summative evaluation will be based on the same criterion. In the case of St. Peter's College, this is largely contained in the Competency and Duties List.

Consistency. Among elements of the approach, it is important that the teacher evaluation for improvement processes are undertaken with consistency. For instance, staff should know that the time lines contained in the written procedures are actually adhered to, that there will be a minimum number of classroom observations, and that expectations will be met concerning all conferences. Moreover, written statements, from performance objectives based on teacher competencies to the final evaluation report, must conform to the agreed-upon formats.

Above all, there must be consistency in the maintenance of an open climate, with mutual respect playing a dominant hand.

Conclusion.

The St. Peter's College application of the Shinkfield model has proved to be a most attractive, successful approach for the evaluation of teachers. Provided that the teacher is being evaluated only with the expectation of positive outcomes in mind, improvement as a professional practitioner of teaching most likely will occur.

The approach also has shown that the teacher appraisal function can be a facilitating and enhancing process, characterized by openness of approach, mutual respect between teacher and principal, and heightened awareness of the importance to teaching of mastery of particular skills and competencies.

A positive climate is the essential basis for successful formative evaluation as a part of a teacher's professional development. Taken together, these elements must strengthen the school itself and, most importantly, student learning.

References

- Center for Research on Educational Accountability and Teacher Evaluation (CRE-ATE). (1991). *Temp A Memo*. Kalamazoo, MI: The Evaluation Center, Western Michigan University.
- Redfern, G. (1980). *Evaluating teachers and administrators: A performance objectives approach*. Boulder, CO: Westview Press.

The National Board for Professional Teaching Standards: Assessing Accomplished Teaching

Introduction

In this chapter we move the emphasis on standards from systems for teacher evaluation to the standards for the assessment of teachers themselves. In this case the focus is placed on experienced teachers who will be given the opportunity to prove the extent of their accomplishment through a national assessment body. This body, the National Board for Professional Teaching Standards (NBPTS), was formed with the primary object of improving the quality of life for U.S. citizens through better and more productive schools. The Board's principal planned means of meeting its objective is **"to establish high and rigorous standards for what teachers should know and be able to do, and to certify teachers who meet these standards"** (1991, p. iii). A further, related goal is to advance appropriate educational reforms that aim to improve student learning.

It has long been recognized, but not acted upon, that excellent teachers too often are unrewarded for the quality of their work. One outcome is that too many fine practitioners leave teaching to the detriment of student learning and ultimately of society itself. Moreover, it appears that the potential skills of fine teachers who remain in schools are underutilized or not used at all. Throughout the 20th century, other professions have established and strengthened their status through national certification systems where high standards have been set and maintained. Such systems for teaching, although proposed in the U.S. periodically, have languished.

The 1983 report of the President's National Commission on Excellence in Education, *A Nation at Risk: The Imperative for Educational Reform*, sharpened public awareness of the sorry state of public education. Few reports about public education have so deeply concerned this country's citizenry. Reform initiatives of many kinds were suggested, many of which resulted in positive action. For instance, in 1986 the Carnegie Task Force on Teaching as a Profession in its report, *A Nation Prepared: Teachers for the 21st Century*, called for the setting up of a National Board for Professional Teaching Standards. In 1987, this bold new venture was born.

Some Aims of the National Board. Designed for teachers with some years of experience, the Board's certification requirements have been planned to focus on ascertaining the extent to which accomplished teachers can use theoretical concepts in practice; have a solid knowledge of subject content; and can preform at an advanced level in teaching/learning situations. The Board's certification is voluntary and is not intended to be mandatory for teacher advancement (although they may serve this purpose), but rather to give recognition to the country's finest teachers as well as offering community assurance of high quality teacher performance.

There is no intention that the Board's certificates will replace each state's system of mandatory licensure. Various states and school districts will decide for themselves how they wish to make use of teachers qualified by the Board to strengthen their schools.

From its outset the Board claimed the commendable objective of hoping to restore public confidence in the schools through rigorous standards for skilled teacher certification. However, whether the National Board's work will eventually redefine teaching as a career remains to be seen. Present planning will see the first significantly large group gain certificates in 1997 (following earlier field testing commencing in 1993). It clearly will be many years thereafter before the cumulative effects are known.

The Board also has held to the contention that its procedures and processes will influence teacher education and teacher professional inservice significantly. Its certificates will reward and recognize teachers who have the ability and initiative beyond licensing requirements and who could become influential in their schools and districts as persons noted for their proven expertise. It should be pointed out, however, that credibility of this national venture will reside in the development and maintenance of an examining system that meets the requirements of standards that are valid and that, in specific terms, have attributes of propriety, utility, feasibility, and accuracy. This will be the crucial test of the true worth of the Board's certification and, therefore, the extent of its influence nationally.

Experienced classroom teachers comprise the majority of the 63-member Board. The reason for this is the belief that they have the expertise to help develop

innovative performance-based assessment methods to measure teacher performance against "standards, the planning and growth of which they themselves are strongly influencing."

Prerequisites for National Board Certification

Much of the Board's examining for certification will be based on "what teachers should know and be able to do." These expectations are laid out well in the 1991 policy booklet, *Toward High and Rigorous Standards for the Teaching Profession* (pp. 13-31). The Board will endeavor to identify and recognize teachers who enhance student learning while demonstrating high levels of "knowledge, skills, dispositions and commitments," reflected through five teacher behaviors presented as core propositions:

1. Teachers are committed to students and their learning.
2. Teachers know the subjects they teach and how to teach those subjects to students.
3. Teachers are responsible for managing and monitoring student learning.
4. Teachers think systematically about their practice and learn from experience.
5. Teachers are members of learning communities.

From each area, the specific skills of accomplished teachers are enumerated and elaborated. This itself is not a particularly onerous task. What is difficult, but completely essential if "rigorous standards" are to apply, is to elicit and depict precise standards as criteria for judgment for each of these five propositions and their extended and operationalized detail. Later discussion indicates that this has been undertaken thoroughly. However, whether or not these activities will result in *valid certification of superior teaching* is a matter for some conjecture.

Who is Eligible for Certification? The NBPTS established prerequisites, or benchmarks, for eligibility to be a candidate for certification requirements. These will act as qualifications or restrictions on eligibility. To this end, the Board has suggested four criteria:

1. Eligibility for Board certification should be as open as possible without compromising essential properties of the standards. (Unexplained is the phrase, "essential properties of the standards." The content of the policy document fails to throw light on this matter.)
2. Any prerequisite adopted should be set in a nonarbitrary manner.

3. Any prerequisite adopted should lend itself to being applied uniformly and fairly to all candidates.
4. Prerequisites should be administratively feasible, lending themselves to straightforward verification.

There are two prerequisites before teachers are eligible to be applicants for the assessment process. Candidates must have at least a baccalaureate degree from an accredited institution and must have "successfully" taught for three years or more at an elementary or secondary school. The Board acknowledges it is embarking on dangerous waters when it includes the word "successful." However, the Board asserts that it "would be recognizing a well-accepted view, that accomplished teaching only comes with practice and time," and then adds that, "This is not the same thing as asserting that the more one teaches, the better one gets" (1991, p. 38). This is consistent with a view that time and practice in teaching are helpful but not sufficient conditions for becoming an accomplished teacher. The Board also emphasizes the voluntary nature of seeking Board certification.

The Board fully realizes the difficulties in simply accepting licensure of any state as a minimum professional requirement, such are the different requirements between states. Considerable diversity does exist between teacher education requirements prescribed by various states. With a quagmire of other difficulties to also consider, the Board's decision to set a 3-year (minimum) teaching period as a precondition for the assessment process seems reasonable and fair.

The Board contends, moreover, that the provision by a candidate of a portfolio of materials developed over time is evidence of experience. Alternatively or additionally, certification requirements could begin to be met during teacher education. These matters are still to be resolved together with what assessment methods will be "administratively feasible, professionally acceptable, publicly credible, legally defensible, and economically viable" (1991, p. 39). The defining of all these terms and their attendant processes will be an essential part of standard setting.

Rationale Underlying National Board Certificates. A fundamental view of the Board is that teaching is content-specific to achieve defined educational objectives. The Board's policy document clearly presents design criteria for certification and allies these policies and procedures to the various subject or learning fields to be examined.

Supporting the fields in which certificates will be given, the Board has selected four criteria to be applied to two dimensions of teaching.

In relation to criteria, issued certificates must

1. support the National Board's standards as expressed in "what teachers should know and be able to do"

2. assure fairness
3. complement the structure of the education system
4. emphasize parsimony

Concerning the first criterion, emphasis will be placed on procedures leading to certification supporting the knowledge, skills, and dispositions stated and implied in "what teachers should know and be able to do." Implications for the assessment exercises are that there are generic knowledge and skills applicable to all teachers and teaching, that there is a distinct pedagogy at particular stages of student development, that there is knowledge of pedagogy that is subject-matter specific, and that teachers must demonstrate depth of content knowledge as well as breadth.

In order to assure fairness, which is the second criterion, the Board intends to avoid discriminating against the chance of any single group of teachers obtaining certification. The policy booklet exemplifies this by stating that "A rural high school teacher who was responsible for math and science instruction should have the same opportunity for certification as a suburban high school science teacher who only teaches physics" (1991, p. 42). Similarly, teachers who work across the traditional boundaries of elementary, middle, or high school should not be discriminated against. Thus, through the design of its certificates and related assessment procedures, the Board will endeavor to treat all candidates as fairly as possible.

The third criterion, complementing the structure of the education system, means that the Board intends to recognize existing good elements in the educational structure while being sufficiently flexible to accommodate changes in that structure. For example, the promising practice of having subject specialists in elementary schools should be recognized by the Board, constructing appropriate assessment procedures for those teachers.

Concerning the fourth criterion, emphasizing parsimony, the Board wishes to make its certification available as soon as possible, to make the certification framework easy to describe, and to keep costs as low as possible. It intends to achieve parsimony in these ways by limiting the dimensions used to define the certificates and by keeping the number of categories in each dimension as few as possible. (Dimensions are described in the next section.) The number of available types of certificates could proliferate unless care is taken. "Since the number of certificates required is a product of the number of categories in each dimension, adding dimensions and categories to the framework creates geometric growth in the number of different certificates to be developed" (1991, p. 42).

The two dimensions of teaching allied to the four criteria are

1. who is taught—the student dimension
2. what is taught—the subject matter dimension.

The first dimension addresses the basically different roles teachers play with students at different grade levels. For instance, subject specialty increases as grade levels rise. As it is well known that the relationship between grade level and stages of student development are not always synonymous, the Board has decided to develop certification fields that overlap the teaching of students at various stages of development and grade levels. This will enable candidates to have a choice of certificates—for instance, early childhood (ages 3-8) or middle childhood (ages 7-12). The Board has decided not to offer certificates in a wide range of specifications, such as teaching related to inner-urban gifted and talented or handicapped students, but rather to encompass a broad range of students reflecting the great diversity of students typically within broad categories of American classrooms. However, for teachers of special needs students, the Board intends to develop assessments for teaching practice specialties that will constitute a further dimension to the certification fields that are outlined shortly (under Policy Decisions on the Framework for Certification). Concerning subject matter dimensions, the Board has endeavored to achieve the criterion of parsimony by offering broad fields, while at the same time has tried to meet the depth of understanding criterion by focusing on a subfield within a major subject-field domain. This suggests that specialty examinations may eventually be necessary, together with demonstrated knowledge in the subject field (e.g., biology within science).

Policy Decisions on the Frameworks for Certification. It has always been the Board's intention to offer certification assessment processes that will become available to all teachers.

Assessment processes leading to certificates will be guided by these policy decisions:

1. There is a core of professional knowledge that all National Board-certified teachers should command.
2. There are knowledge, skills, and methods particular to different stages of student development that teachers working with certain students should command.
3. There are subject and discipline-specific knowledge, skills, and methods that teachers should command, including a core of subject knowledge and discipline-specific knowledge teachers in each subject area should command.
4. Each certificate will be designed to require a demonstration of depth as well as breadth of knowledge.

Moreover, the Board intends to modify certificates as time passes if appropriate reviews indicate that this should happen.

The Board intends to focus research and development activities in the following areas:

- Early Childhood (Ages 3 - 8)
 - Generalist
- Middle Childhood (Ages 7 - 12)
 - Generalist
 - English/Language Arts
 - Mathematics
 - Science
 - Social Studies/History
- Early and Middle Childhood (Ages 3 - 12)
 - Art
 - Foreign Language—Spanish, French, and others
 - Guidance Counseling
 - Library/Media
 - Music
 - Physical Education/Health
- Early Adolescence (Ages 11 - 15)
 - Generalist
 - English/Language Arts
 - Mathematics
 - Science
 - Social Studies
 - Adolescence and Young Adulthood (Ages 14 -18+)
- English/Language Arts
 - Mathematics
 - Science
 - Social Studies/History
- Early Adolescence through Young Adulthood (Ages 11 -18+)
 - Art
 - Foreign Language—Spanish, French, and others
 - Guidance Counseling
 - Library/Media
 - Music
 - Physical Education/Health

— Vocational Education—Agriculture, Business, Health Occupations,
Home Economics, and Industry/Technology

In brief, successful candidates will need to demonstrate a strong knowledge of four aspects of education:

1. Core professional knowledge (including material on development, cultural and linguistic diversity, classroom management, and the history of schooling in U.S. society).
2. Developmental specific knowledge (including in-depth knowledge of human development to the appropriate student level, and also the application of that knowledge to instructional settings).
3. Breadth of content and discipline area knowledge (including understanding and appreciation of subject matter and related pedagogical expertise).
4. Depth of content and discipline area knowledge. (This would be a subset of the main certificate field, as explained earlier.)

While these four strands have been isolated for definition purposes, they may very well be combined in simulated exercises that form part of the assessment procedures. Within the six developmental levels listed, both knowledge and school requirements will be essential. The Board is presently undertaking more extensive consultations with relevant subject-matter groups; evaluating the knowledge base for each field; and researching the most promising alternatives "for assessing breadth and depth of knowledge with respect to reliability, validity, and efficiency" (1991 p. 51). If this last activity is carried out scrupulously, the public's and the teaching profession's confidence in the subject-related examinations will be considerably enhanced. It would constitute, in this particular aspect of the National Board's intentions, sound standard setting.

Principles Guiding Development of Assessment

Apart from the Board's vision of "what teachers should know and be able to do," the Board also wishes to envisage teaching "as a collegial enterprise involving complex decision-making" (1991, p. 53). Since the Board's aim is to place these (and other) concepts in a context that is professionally credible, publicly acceptable, legally defensible, administratively feasible, and economically affordable, it has issued a series of policies related to the assessment development process.

In brief (inter alia) the Board expects

- assessments to measure what teachers should know and be able to do to help student learning (the importance of this as a validity issue will be discussed later)
- assessment procedures to “profoundly” affect the teacher’s role in education
- the assessments to offer a variety of methods, including those that are most up-to-date
- the assessments to be affordable and accessible to all experienced teachers
- professional subject and other associations to be actively involved in the process
- to work collaboratively with appropriate state agencies and research and development centers in the development of the assessment process
- to eliminate all prejudicial biases in the development process
- the assessment process to provide useful information for teachers as well as constructive feedback (This feedback policy aim may not be realized as later discussion will indicate.)

Planned Methods for Assessments

The National Board has considered a wide range of assessment methodologies, including multiple choice essays, interviews, simulated contexts (with questions taking the form of multiple choice essays and interviews), simulated performances (e.g., of teaching practices), documentation (portfolios and videotapes related to a candidate’s school-site situation), limited observation by trained assessors, and regular on-site observations by others.

The Board intends to make a choice among these options; however, the policy of offering a variety of assessment approaches will be adhered to. A decision was reached early on that assessment centers will be required, with their attendant expert administrators and assessors. Although the Board has stated the intention of directing research toward methodologies that might assess a teacher’s capacity to integrate knowledge from different sources and to participate in school decision making, this goal has proven very difficult to plan in specific detail.

The Board assumes that some aspects of subject matter and generic pedagogical knowledge may be assessed through essay or multiple choice methods, but that the attainment of standards relating to actual teaching practice (including, presumably, effects on student learning) and professional judgment will require simulation, observation, and documentation. Of these last three mentioned methods, observation, many believe, is the critical factor. Indeed, expert evaluators have stated that the validity and therefore the credibility of the Board’s work will rise and fall to the extent that observation is professionally and effectively carried out. This matter is discussed later (under Summary on Standards and Validation).

Criteria for Selecting a Particular Methodology The National Board intends to use three criteria in the selection of assessment methodologies. These will be examined during the research and development process:

1. Validity—the extent to which the assessment procedures measure the standard
2. Efficiency—the feasibility of the process in terms of time and money and quality of derived information
3. Impact—the extent to which the whole process strengthens teaching practice nationally

In line with the Board's aim to meet these criteria, a Technical Analysis Group (TAG) was instituted at the Center for Educational Research and Evaluation, University of North Carolina at Greensboro. This group subsequently requested The Evaluation Center at Western Michigan University to prepare a guide for the Board's Assessment Development Laboratories (ADLs). This guide has taken the form of an evaluation criteria framework for teacher assessment systems. TAG's work, for all its formidable problems, appears to be undertaken with commendable professionalism.

Education Policies and Reform Priorities. The principal aim of the National Board is to improve student learning by strengthening the teaching profession. A delicate point arises when it is considered that educational policy and practice, constitutionally, has always been a state prerogative. Although the Board states that it wishes to work within the existing framework, state endorsement for what it plans to do has not been universal. It realizes that future acceptance will depend very strongly on the nature of its standards and assessments policies, procedures, and outcomes. The Board intends to play a variety of collaborative roles to enhance its acceptance nationally. It hopes that by defining high and rigorous teaching standards and by applying these to its assessment procedures, a national teaching force of superior and recognized quality will evolve. Three reform issues have been selected as the focus for efforts to improve teaching and learning:

1. creating a more effective teaching and learning environment in schools
2. increasing the supply of high-quality entrants to the profession
3. improving teacher education and ongoing professional development

Concerning the first issue, the Board wishes to define the characteristics of a professional teaching workplace, to promote these principles, and to identify schools that exemplify progress toward meeting the expressed ideals. In relation to increasing the quality of teacher entrants, the Board intends to collaborate with others whose support seems most appropriate (e.g., educational institutions and

teachers themselves) to begin to make improvements. It is especially important that gifted minority students are encouraged to consider teaching as a career. Concerning improving teacher education and professional development, the Board intends to communicate widely with teachers (at all levels) and policymakers (federal, state, and local) to obtain reactions to Board standards, particularly to the ways in which higher education can better prepare intending teachers to meet those standards.

Planning and Progress

From its inception, the NBPTS planned a number of main objectives to be completed by (approximately) 1993.

A strategic plan outlined the steps to be followed to complete four of these objectives:

1. To identify the elements of accomplished teaching and to convert these to high and rigorous standards for the Board's assessment system
2. To develop the best possible assessment methods to meet the prescribed standards
3. To promote educational policy and reform to improve teaching and learning
4. To become a self-financing, nonprofit organization with strong communication and marketing skills

A second phase of planning involved research and development, ongoing processes that will support all future developments of the Board's endeavors.

Certification Standards. A major undertaking of research has been the development of standards planned for each certification field listed earlier. This basically entails converting the underlying attributes of "what teachers should know and be able to do" into curriculum which, taken together with the assessment system (for each certification field) by ADLs, constitutes a standard. This has and will involve a close collaboration of teachers, other educators, researchers, and selected lay leaders. By 1994 six initial drafts of certification area standards had been prepared for field testing and other uses, including the development of assessment systems, all of which are planned to lead to their refinement and compatibility across assessment fields.

The Board anticipated that the development of these standards would prove to be a difficult task, and indeed it has been so. Some early standards, it was conjectured, would not yield valid or affordable assessments. Two standards committees and their related assessment systems, however, had their work sufficiently and satisfactorily reviewed to be used for the first field tests in the latter part

of 1993. They are the Early Adolescence/Generalist and Early Adolescence English/Language Arts standards.

In brief, these early assessment area standards committees consider the five core propositions of "what teachers should know and be able to do"; take these to the next level of specificity, that is, describing what they mean in a particular field; translate this description into curriculum; and have ADLs complete the process by developing appropriate assessment procedures. These activities develop standards, and there is a three-fold way of describing each:

1. Definition of the standard
2. Elaboration
3. Vignettes and commentary

Even in draft form these standards documents display a thorough, professional, and convincing approach. They undoubtedly will become recognized as solid standards for aspects of those types of assessment planned to distinguish accomplished teachers. Whether or not those assessments in themselves are sufficient to complete the task is another matter altogether.

Researching Assessment Product Development. The most costly and enterprising aspect to the Board's work has been, and will continue to be, the actual development of assessment systems that are valid and acceptable. It is estimated that \$50 million will be needed to complete this task. To this end, the Board has commissioned a series of Assessment Development Laboratories (ADLs) that will work closely with teachers and other selected personnel to develop both instruments and procedures.

In addition to these ADLs, support for the Board's mission will come from cross-cutting research (of the type outlined earlier when the evaluation criteria framework was discussed), a Technical Analysis Group (also referred to earlier), and a field test network (whose work commenced in 1993).

The Technical Analysis Group (TAG) can be described as the research arm of the Board. Among other activities TAG is continuing to pursue a series of studies of the psychometric quality and defensibility of the Board's teacher assessment packages. For instance, one validation study is examining the capability of the NBPTS assessment packages for Early Adolescence/Generalist and Early Adolescence English/Language Arts fields to discriminate between teachers who differ substantially in levels of teaching expertise. A series of positive findings would support the claim that these assessment standards can be used to identify teachers who are accomplished.

Implementation: Field Testing. Although the Board's published intention was to begin assessing its first group of candidates for certification in 1993, the first field test was not completed until early 1994. Whether the teachers who were subjects of the field tests will gain certification, if assessed satisfactorily, is a matter for later Board decision. The Board is aware that the awarding of certificates based on imperfect standard-setting and assessment procedures could be very damaging to its avowed aims and ideals.

As mentioned, the initial field test focused on two NBPTS certificates, the packages for which have been prepared by two ADLs. These packages were field tested on a self-selected nationwide sample of teachers. During 1994 they conducted a number of studies involving assessment center coordinators and field test network coordinators in an endeavor to ascertain the psychometric quality of the assessments, using data collected during the field tests. If the assessment packages are judged by TAG to be valid, reliable, and free from bias, the National Board may certify qualified candidates in the fall of 1994.

Summary Discussion of Standards and Validation

It is possible to identify strengths and potential weaknesses related to the Board's aim to acknowledge "successful" teachers with their accolade. Some of the more obvious are briefly discussed.

Strengths of the Proposed System. The following appear to be six very commendable features of the Board's approach for high and rigorous standards for the teaching profession:

1. The Board has sought communication and collaboration with a wide range of persons, associations, and other organizations vital to the success of this national venture.
2. The Board has endeavored to define the core elements of "what teachers should know and be able to do."
3. Resulting from 2 (above) further defining of teaching levels, generic teaching qualities, and subject areas has been well developed.
4. The explication of standards undergirding the credibility of areas to be assessed has commenced, with strong representation from expert teachers and other relevant persons.
5. The research and development work has been well planned, particularly in respect to the involvement of the ADLs and the encompassing functions of the TAG (including, importantly, formative evaluation of field test procedures).

6. The Board's basic aim to improve teaching and learning nationally and, as a result, to influence the quality of educational provisions more widely, in combination, are most worthwhile objectives.

Potential Weaknesses of the Proposed System. The following 5 points can be construed as real or potential weaknesses in the Board's work:

1. The chief concern of the authors of this book and other evaluators relates to the general credibility of the assessment system, and particularly the assessment of a "successful" teacher. Although the Board intended that candidates would be observed in actual classroom situations, this has not occurred in field trials. Trying to draw effective teaching correlations and credible conclusions from videotapes, assimilated exercises, and written examinations too easily can be a distant cry from classroom realities, including, most importantly, student learning. In fact, these activities are close to quasi evaluation, no matter how rigorous developed standards are for assessment fields. In short, unless there is a thorough, well-planned, well-executed observation of candidates by credible evaluators, the validity of outcomes is highly suspect.
This need only be a potential weakness, since the TAG is well aware of the importance of expert classroom observation giving credibility to certification. Research into this problem has already begun. It is hoped that this matter will be satisfactorily resolved before the certificate is offered in its full range of assessment fields.
2. An associated weakness, one that also relates to the definition of a "successful" teacher, is that the Board's procedures beg the question: "Has students' learning been identified?" Videos, no matter how refined and further refined, will *not* indicate this. As one of their essential duties, teachers must aid student learning, and this must be demonstrated. For all their "state-of-the-art" sophistication, the assessment methodologies do not address actual, perceived, and recorded student learning. It is too big a leap of faith to assume that these methodologies, which clearly will measure the extent of some commendable teacher attributes, will measure *all* attributes. If a basic attribute, like improving student learning, is not measured, the Board certificate is not necessarily distinguishing the "successful" teacher. This major validity (and reliability) problem must be addressed.
3. The Board states that an experienced teacher should have the benefit of "collegial cooperation" as part of his or her professional development. While no one would argue against this, the Board would need to prove the long-term validity and efficacy of any collaborative approaches to teaching

(and student learning) and methods of discerning accurately the input of various teachers into the total product.

4. The Board's standards, which are being developed independently by ADLs, run the risk of lacking sufficient uniformity to give comparative equity and credibility of assessment fields. Put another way, will one certificate be perceived to be better/different/more difficult to obtain than another? The TAG is aware of this potential trouble spot, together with the allied problem of balancing specificity of subject matter (in an assessment field) with the stated general aims of the Board.
5. While the assessments might reveal a rich array of teacher capabilities (and, indeed, the field tests have done this), scoring could prove a very real difficulty in such a major enterprise. Unless scoring is both valid and reliable, confidence in the entire scheme cannot be sustained.

Benefits and Costs of NBPTS Outcomes

Potential clients of the Board's work include teachers interested in gaining this superior certification, states, school districts, and schools, which may decide to support teachers by meeting the cost of the certification process. Thus, even in advance of certificates being awarded, we consider it useful for consumers to be aware of possible benefits and costs.

Benefits. Real and lasting benefits of the Board's work will happen only if the certificate is strongly accepted nationally. It is, of course, the Board's intention that this will occur and, indeed, all of its efforts have been directed toward achieving this ideal, which is expressed in the NBPTS mission: to establish high and rigorous standards for what teachers should know and be able to do, and to certify teachers to meet these standards. Assuming that national recognition eventuates, with the majority of teachers seeking certification at some stage of their career, these benefits should occur:

- The status of teachers as professionals will be enhanced.
- Individual teachers will be considerably helped toward having their professional ambitions realized (in areas such as satisfaction of professional achievement, promotion within a present system or one further afield, or reassignment to a chosen pedagogical field).
- State and district policies that enhance teacher mobility will be instituted.
- Individual schools or school districts will have their prestige enhanced if a considerable number of their teachers (perhaps the majority) gain certification.

- School districts could use the Board's certificate as one significant measure in a systemwide teacher evaluation process.

Costs and Risks. There seems little doubt that the main risk associated with the Board's work rests squarely with troubled inner city schools and school districts and, to a lesser extent, small rural districts. Both groups could see many of their best teachers creamed off by wealthy suburban schools and school districts. The potential costs to these poorer schools, over a period of time, could be enormous. Such school districts may, of course, decide to make use of the Board's certification and its standards by encouraging teachers to gain certification and by endeavoring to retain them through a concerted effort to gain community support for higher teacher salaries.

Other costs do not assume the same proportions as losing accomplished teachers. They could, however, include

- School districts may give too great a weighting to the Board's certificate by comparison with other assessments of a teacher's merit.
- The financial burden to a poorer school district could be considerable if it is decided that most of its teachers should be supported in the quest for NBPTS certification.
- Within the staff of a particular school (or school district), animosity may arise if it is felt that disproportionate favors are going to Board-certified teachers.
- Since preparation for assessment may involve considerable time and effort on the part of the teacher who chooses to participate in the Board's assessment, teaching in the participating schools may suffer.

Early Leads From States and Districts

In an *Education Week* article (September 7, 1994, p. 14) entitled "States Offer Incentives to Teachers Seeking National Board Certification," Joanne Richardson, citing the NBPTS as the source, gives details of some states and districts that are offering financial and other incentives for teachers seeking National Board certification. The fact that some states have approved measures to support teachers in various ways and that other states are contemplating doing so is a sign that the *national* intention of the Board's work may be realized.

North Carolina, prompted by recommendations from a panel convened by the governor, has passed the most comprehensive legislation. The legislature has allowed \$500,000 in 1995 to cover costs of assessment fees (presently \$975 per teacher) and intends to give a 4 percent salary raise to successful candidates. In addition teachers will be allowed up to 3 days of release time to prepare portfolios

and other assessment related activities. The state has also adopted a policy that will permit out-of-state teachers to practice in North Carolina without meeting state-specified requirements.

New Mexico has set aside \$315,000 to support teachers in their preparation for Board certification. Teachers who relocate to Oklahoma will have state certification waived if they gain a National Board certificate. Moreover, the state's Board of Education is considering an incentive scheme to encourage teachers to put themselves forward for national certification. In Iowa, state funds for staff development may be used for National Board Assessments, and successful teachers will automatically receive a state license if state officials are convinced that the Board's standards meet or exceed the state's. Teachers in Mississippi who are National Board certified will receive a \$3,000 bonus when 80 percent of the Board's proposed 30 assessment areas become available.

Richardson's article also reports that several districts "have approved policies as teachers' contracts to complement the Board's work." (p. 14). For example, the Boston and Vancouver, Washington, districts have set policies encouraging teachers to gain Board certificates. The Boston situation is particularly interesting, as a contract was recently negotiated between the Boston teachers' union and the school district that will reimburse 25 teachers a year for Board fees and make teachers who successfully complete assessments eligible for "lead teacher" status. This title gives both a 10 percent salary increase and additional professional duties. Teachers in the Vancouver district who become Board certified are also to be given financial rewards and special status. Moreover, an allowance of \$500 will be offered to all teachers who present themselves for Board certification.

The incentives offered or planned by these states and districts (and others) may suggest that political forces are beginning to realize the potential worth for states, districts, and teachers of Board certification. If a national award has helped to raise the status of other professions, it is reasonable to expect that teachers could similarly benefit.

Conclusion

The National Board's aim to raise the status of accomplished teachers is most worthwhile. If this goal is reached, there will be obvious benefits for important aspects of education nationally. There are already indications from the field testing and other activities involving teachers that the preparation for the assessment procedures is a valuable professional development for teachers. If it does eventuate that a legally defensible process can be developed for giving teachers feedback, the experience will be even more valuable.

Present research projects to address real and potential weaknesses may very well determine whether the National Board's goals will be attained. The fact that these weaknesses have been realized some years before certificates will be widely offered is a positive situation for the Board.

Lee Shulman, Professor of Education at Stanford University, summed up the bold stance that has been adopted by the Board in a recent interview (March 1994) with an *NEA Today* writer:

The teaching profession has to appreciate what an extraordinary experiment the National Board is engaged in. If we're realistic, we'll expect serious problems. If there are only a few, the operation will be incredibly successful. No one's ever tried to do this much all at once.

References

- Jaeger, R. M. (1993, June.) *The current measurement research agenda of the technical analysis group to the National Board for Professional Teaching Standards*. Presentation at the CREATE National Institute on Evaluation. Kalamazoo, MI: Center for Research on Educational Accountability and Teacher Evaluation.
- National Board for Professional Teaching Standards. (1991). *Toward high and rigorous standards for the teaching profession*, Third Ed. Detroit, MI: Author.
- National Board for Professional Teaching Standards. (1994). Various information/promotion documents and pamphlets. Detroit, MI: Author.
- Richardson, J. (1994, September 7). States offer incentives to teachers seeking National Board certification. *Education Week*.

The Tennessee Value-Added Assessment System (TVAAS): Mixed Model Methodology in Educational Assessment.

By William L. Sanders³ and Sandra P. Horn⁴

Abstract

The Tennessee Value-Added Assessment System (TVAAS), developed by Dr. William L. Sanders and associates, is a method of assessing the impact of educational systems, schools, and teachers on the gains students make from year to year on norm-referenced achievement tests. By collecting and aggregating data on students and teachers over several years and employing mixed-model statistical methodology, TVAAS can provide unbiased measures of the influence of school systems, schools, and teachers on student academic progress.

Introduction

Background. Over the past decade, Tennessee, like so many other states, has continuously sought to improve educational opportunities for its students. The first wave of reform resulted in the Comprehensive Education Reform Act of 1983 (CERA). CERA created a Career Ladder Program (a merit pay system for teachers) and a Basic Skills Program. CERA also led to the articulation of grade and subject curricula and the development of curricular frameworks for the state of Tennessee.

At about this time, independent of the efforts of the Tennessee Department of Education, two statisticians, Dr. William L. Sanders and Dr. Robert A. McLean of the University of Tennessee, had begun to explore the feasibility of using statistical mixed model methodology to eliminate many of the previously cited impediments to incorporating student achievement data in an educational outcome-based assessment system. These problems include but are not limited to the following: missing student records, various modes of teaching (i.e., self-contained classroom vs. departmentalized instruction vs. team teaching), teachers changing assignments over years, transient students, regression to the mean, different variance-covariance structures across school systems, and the need to include concomitant covariables

³ William L. Sanders is Professor and Statistician, Agricultural Experiment Station, Statistical and Computing Services, University of Tennessee.

⁴ Sandra P. Horn is an educational consultant and a media specialist with the Knox County Schools.

as needed. A decade of work has demonstrated that a system can be developed to eliminate, or at least trivialize, these problems.

In 1984, McLean and Sanders published a working paper on the use of student achievement data as a basis for teacher assessment. Utilizing three years of gain scores from Knox County students' performance on the California Achievement Test in grades 2 through 5, Sanders and McLean developed a statistical system of analysis based upon Henderson's mixed-model methodology. This study rendered the following findings:

1. There were measurable differences among schools and teachers with regard to their effect on indicators of student learning.
2. The estimates of school and teacher effects tended to be consistent from year to year.
3. Teacher effects were not site specific, i.e., a gain score could not be predicted by simply knowing the location of the school.
4. There was very strong correlation between teacher effects as determined by the data and subjective evaluations by supervisors.
5. Student gains were not related to the ability or achievement levels of the students when they entered the classroom.

Subsequent studies incorporating data from Blount County and Chattanooga City Schools bore out the initial findings. The study of the Chattanooga City Schools, a system that includes many inner-city schools, produced a new finding not evident from the previous studies of systems that were primarily suburban and rural: the estimate of school effects was not related to the racial composition of the student body.

Even though these findings indicated the efficacy and utility of this assessment approach, the Sanders model (as this process has been labeled in Tennessee) was for several years thereafter known only to a small circle of educators and statisticians.

In 1988, educational reform in the state took a different direction. The Tennessee Department of Education developed a document titled 21st Century Challenge: State Goals and Objectives for Educational Excellence in response to the America 2000 Program. The Tennessee State Board of Education put forth its Master Plan for Tennessee Schools and the Tennessee Higher Education Commission developed Tennessee Challenge 2000 for postsecondary educational institutions. The goals and objectives of these governing bodies were coordinated to form an educational framework that would address learner needs and expectations from preschool through adulthood. At every level, the need for accountability and assessment was recognized as an essential component of educational improvement.

When the recommendations of the governing educational bodies were submitted to the Tennessee General Assembly for legislative action in the form of the Education Improvement Act, it became necessary to specify the means by which teachers, schools, and school systems would be held accountable for meeting the goals and objectives set forth for Tennessee's educational systems. Since the focus of the accountability movement was on the **product** of the educational experience rather than the **process** by which it was to be achieved, the outcomes-based assessment system Sanders and McLean had been refining was an obvious choice for consideration. In 1991 when the Education Improvement Act was adopted, the model now known as the Tennessee Value-Added Assessment System (TVAAS) formed an integral part of the legislation.

Philosophical Underpinnings of the Tennessee Value-Added Assessment System.

Ralph W. Tyler, a major force behind the development of modern educational evaluation, proposed that evaluation should be a process of comparison between stated objectives and actual outcomes. In Tennessee, the connection between objectives and outcomes is explicitly recognized. *The Master Plan for Tennessee Schools 1993* sets forth goals in eight key result areas: early childhood education; primary and middle grades education; high school education; technology; professional development and teacher education; accountability; school leadership and school-based decision making; and funding. The goal for the accountability component of the master plan is as follows: "State and local education policies will be focused on results; Tennessee will have assessment and management information systems that provide information on students, schools, and school systems to improve learning and assist policy making." (Tennessee State Board of Education, 1992, p. 7). Here, Tyler's conception of evaluation is readily discerned. Assessment is recognized as a tool for educational improvement, providing information that allows educators to determine which practices result in desired outcomes and which do not. By focusing on outcomes rather than the processes by which they are achieved, teachers and schools are free to use whatever methods prove practical in achieving student academic progress. Value-added assessment is one means recognized by the state of Tennessee for assessing progress toward the academic goals set forth in the master plan (p. 17) and the Education Improvement Act.

Astin (1982, p. 14) states that "the basic argument underlying the value-added approach is that true excellence resides in the ability of the school or college to affect its students favorably, to enhance their intellectual development, and to make a positive difference in their lives." TVAAS was developed on the premise that society has a right to expect that schools will provide students with the opportunity for academic gain regardless of the level at which the students enter the educational venue. In other words, all students can and should learn commensurate with their abilities. By focusing on the gains that all students make from year to year, the

school systems and the individual schools deemed to be most effective by TVAAS are those that provide educational opportunities for all learners—the advanced learner as well as the slower learner.

A Description of the Tennessee Value-Added Assessment System

General Information. TVAAS is a statistical process that provides measures of the influence that school systems, schools, and teachers have on indicators of student learning. Initially, TVAAS will furnish this information on the system level for each school system in Tennessee for grades 3 through 8 in math, science, reading, language, and social studies by using the scale scores from the Tennessee Comprehensive Assessment Program (TCAP). TVAAS will be extended to cover grades 9 through 12 when subject matter specific tests that can provide comparable data for these grades have been developed and validated—by law, no later than July 1, 1999. TVAAS is mandated by the Education Improvement Act that took effect July 1, 1992.

TVAAS analyzes the scale scores students make on the norm-referenced items of the TCAP. The pattern of the scale scores over the child's school career forms a profile of academic growth. A database containing the merged records of all students in Tennessee who have taken the TCAP tests during the past 3 years has been constructed. At present, it contains more than 1.6 million student records. This number will continue to grow over time and will enable continued tracking of the academic growth of each student.

The Education Improvement Act (EIA) mandates that school system effects on the educational progress of students for grades 3 through 8, as determined through the use of TVAAS, will be reported for systems statewide no later than April 1, 1993. These reports have been distributed to each school system in Tennessee. They have also been released to the public and will be updated annually.

The EIA sets July 1, 1994, as the deadline for issuing the first set of reports on individual school effects. This set of reports will also be available to the public and will be revised on a yearly basis.

The individual teacher effects for teachers of grades 3 through 8 are to be reported to the teacher, appropriate administrators, and school board members no later than July 1, 1995, according to the EIA. These reports relating to the influence of individual teachers on the rate of student learning will not be available to the public. Reports on all levels will be based on at least three years of data and no more than five years of data.

The Assessment of Schools and School Systems. The assessment of schools and systems, although it requires massive computing capabilities, logistical plan-

ning, and statewide testing, is fairly simply explained. The mixed-model equations incorporate the scale scores of all the students taking the norm-referenced portion of the TCAP in all five subjects, modeling a learning profile of each student for each subject as explained in the section above. These profiles are grouped by system or school, as the case may be. The gain scores of a school or system's students are estimated and are then compared to the national norms. Deviation from the norm gain is reported for each subject and grade. The school or school system can then identify where students are achieving normally, outstandingly, and substandardly.

Tennessee monitors the gains of all school systems in the state for subjects or grades that are not achieving national norm gains. Those systems achieving two or more standard errors below the national norms must show positive progress or risk intervention by the state. Each school and system is expected to achieve the national norm gains regardless of whether its scale scores are above or below the national norm.

Assessment of Teachers. The assessment of teachers is generally the most controversial aspect of any educational evaluation system. The great variety of teaching situations and the endless diversity of the student population have rendered each attempt at teacher assessment suspect to a greater or lesser degree. TVAAS is not the first to base teacher assessment on student achievement. However, important differences exist between TVAAS and its predecessors.

Beginning with the 1992-93 school year, detailed information identifying each teacher with the students s/he teaches will be collected annually. Included in the data will be subjects taught to each student and the proportion of time each student spends with a teacher. If team teaching or departmentalized teaching takes place, it will be identified along with the proportion of each subject the teacher is responsible for teaching. From attendance records submitted to the state, it will be determined whether each student has been present in each teacher's class the required 150 days in a given school year, because students who have not been in a teacher's class at least 150 days in a year will not figure into the teacher's assessment. By 1995 when the teacher assessments are scheduled for delivery, 3 years of such data will be available. The EIA requires that teacher assessment be based on at least 3 and no more than 5 years of data.

Test Reliability and Relevance. TVAAS uses scale scores from the norm-referenced items on the Tennessee Comprehensive Assessment Program (TCAP), which was first implemented in the 1989-90 school year. The norm-referenced part of TCAP, the CTBS/4, is a nationally normed test mandated in Tennessee for grades 2 through 8 and grade 10. It assesses skills in reading, language arts, math, science, and social studies. The norms for the test were established in 1989. Williams (1989) states in his review of customized standardized tests that in Tennessee, "the

norm-referenced module was specifically created so that it has proper statistical characteristics of reliability, adequate floors and ceilings, and articulation across test levels." To insure test validity, the EIA mandates that "fresh, non-redundant tests" be used each year. This means that only a small percentage of the items on the CTBS/4 can be carried over from one year to the next. Moreover, rigorous sanctions are provided in the EIA for any breach of test security. The relevance of the test to Tennessee's academic program may be inferred from the tendency of scores across the state to approximate or slightly exceed the national norms in all subject areas and all grades.

The scores from the CTBS/4 cannot reflect the totality of a student's learning experience or progress. However, these scores, as they are utilized by TVAAS, provide an unbiased estimate of the influence of school systems, schools, and teachers on students' academic growth in the subjects tested. This academic growth is and should be a primary goal of Tennessee's educational system. TVAAS uses data from a testing system already mandated and in place statewide. However, should better tests be developed in the future, no major alterations would have to be made in order for TVAAS to incorporate new sources of data, as long as the methods of assessment can provide linear metrics.

Problems of Using Student Achievement Data in Educational Assessment.

The use of student achievement data to directly measure educational outcome has much intuitive appeal and has been advocated by many. However, serious proponents of this approach have recognized several difficulties that must be overcome in order to assure a fair and reliable system for outcome assessment.

These difficulties may be categorized into problems associated with (1) the definition and construction of appropriate metrics and (2) development and implementation of a statistical methodology that will allow fair and unbiased assessment of school systems, schools, and teachers when nonrandom assignment of students is assured. (We will not deal with the definition and construction of metrics, but rather will assume that metrics exist or can be constructed that adequately proxy learning.)

Even when metrics exist with suitable characteristics, many problems of school and teacher assessment remain. The ensuing discussion will focus on the problems associated with the estimation of the influence of **teachers** on the rate of student gain, because at the classroom level the problems are more difficult than at the school or school system level.

Since random assignment of students to teachers is usually not practiced and seldom is possible, simple means of class achievement test scores are seriously

biased by many factors other than teacher influences that affect student learning. Travers (in Millman, 1981) listed (1) teacher influences, (2) parental influences, (3) genetic endowment, (4) other school influences, and (5) availability of materials as being some of the most important factors that determine the rate of student learning.

Later, in their attempt to develop a value-added method of evaluation based on student test scores, Bingham, Heywood, and White (1991) list 44 variables under 5 major categories—individual characteristics; family characteristics; classroom characteristics; school characteristics; and academic performance—which they determined were independent of the input of school and teacher for the subject school system during the years of their study.

In spite of the detailed character of this listing, Bingham et al. point out that these variables may be pertinent only to the particular school system they studied and perhaps only during the years in which the research took place (pp. 200-201). Obviously, any system that will fairly and reliably assess the influence of teachers on student learning must partition teacher effects from these and other factors. However, it is a hopeless impossibility for any school system to have all the data for each child in appropriate form to filter all of these confounding influences via traditional statistical analysis.

Using a different approach, the three studies conducted by Sanders indicate that these influences can be filtered without having to have direct measures of all of the concomitant variables. By focusing upon measures of academic gain, each student serves as his or her own "control" or, in other words, each child can be thought of as a "blocking factor" that enables the estimation of school system, school, and teacher effects upon the academic gain with the need for few, if any, of the exogenous variables.

In an attempt to partition the teacher and school effects from the partial confounding with class ability level, the well-known linear model techniques of analysis of covariance and ordinary multiple regression have been suggested by Millman (1981) and others. The obvious intent was to adjust differences that exist among students to enable a fairer evaluation of teachers. However, if these simple approaches are applied and even if all of the concomitant data were available, still unanswered is the well-known problem of regression to the mean of the teacher effects, which would provide unfair rankings of teachers with varying quantities of student achievement records. Also, the problem of missing student records due to transient student populations, students being absent during the time of testing, and so on would result in very few usable records if these traditional methods were employed.

Advantages of Considering Educational Outcome Assessment from Student Data as a Statistical Mixed Model Problem

Traditional multiple regression or analysis of covariance can be characterized as techniques in linear model analysis with all fixed variables. If the problem is viewed not as a fixed effects problem but rather as a mixed model problem with both fixed and random effects, then much theory and methodology exist that offer solutions to many of the problems that have been cited as reasons for not doing educational outcome assessment from student achievement data.

General Form of Henderson's Mixed Model Equations (MME)

$$y = XB + ZU + e$$

where

y in the context of teacher evaluation is the $m \times 1$ observation vector representing all of the scale scores for individual students for all academic subjects tested over all grades.

X is a known $m \times p$ matrix.

B is an unknown, $p \times 1$ vector of fixed effects.

Z is an $m \times q$ incidence matrix.

U is an unobservable $q \times 1$ random vector.

e is an $m \times 1$ vector with $E(e)=0$.

Both **U** and **e** have null means and variance.

$$\text{Var} \begin{bmatrix} U \\ e \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}$$

G and **R** are known and nonsingular. **R** is the variance-covariance matrix that reflects the correlation among student scores within teacher. **G** is the variance-covariance matrix that reflects the correlation among teacher effects (both **R** and **G** are assumed block diagonal in the context of teacher evaluation). If (**U**,**e**) are normally distributed, the joint density of (**y**,**U**) is maximized for variations in **B** and **U** by the solution to the following equations:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

Let a generalized inverse of the coefficient matrix be

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

Some of the properties of a solution of these equations are as follows (Henderson, 1984):

1. $K'b$ is best linear unbiased estimate (BLUE) of the set of estimable linear functions, $K'B$.
2. u is the best linear unbiased predictor (BLUP) of U .
 - $E(Uu) = u$.
 - $\text{var}(u-U) = C_{22}$.
 - $\text{var}(K'b + M'u - K'B - M'U) = (K'M')C(K'M)'$.
 - u is unique regardless of the rank of the coefficient matrix.
3. $K'b + M'u$ is BLUP of $K'B + M'U$ provided $K'B$ is estimable.
4. With G and R known, the solution is equivalent to generalized least squares and if u and e are multivariate normal then the solution is maximum likelihood.
5. If G and R are not known, then as an estimated G and R approach the true G and R , the solution approaches the maximum likelihood solution.
6. If u and e are not multivariate normal, then the solution to the MME still provides the maximum correlation between U and u .

For an introduction to Henderson's Mixed Model Methodology, see McLean, Sanders, and Stroup (1991).

Why Should Teacher Effects Be Considered Random Instead of Fixed?

Historically, classification variables in a linear model context that have their own probability distribution have been referred to as random effects. Since in the context of teacher evaluation other variables that do not have their own distribution (fixed effects) sometimes may be included to insure fair evaluation, it is often more reasonable to view teacher evaluation as a mixed model problem. When this is the case, solutions to Henderson's mixed model equations (MME) provide Best Linear Unbiased Prediction (BLUP) of the random effects while providing opportunity for the inclusion of both continuous and classification fixed effects. This is a sufficient

procedure to provide the flexibility necessary to handle the diversity of models that could be encountered in teacher assessment. Additionally, since BLUP is a "shrinkage" estimate of the realized value of the random variable (Harville, 1976), then BLUP is a solution to the regression to the mean problem, which has been long recognized as an impediment to the use of student data in an assessment system for teaching effectiveness.

Concept of Best Linear Unbiased Prediction (BLUP). To illustrate the concept of best linear unbiased prediction, a restatement of an example presented by Henderson (1973) from Mood (1950) is presented:

Given that the population mean and variance of true IQ is 100 and 225, respectively, and if an individual takes one IQ test and scores 130 on a test that has test error variance of 25, what is the best prediction of true IQ of that individual? In this example,

$$\begin{aligned} \text{Prediction of true IQ} &= \text{Pop. Mean} + \left(\text{Mean of IQ Tests} - \text{Pop. Mean} \right) \times \left(\frac{\text{Var(Pop.)}}{\text{Var(Pop.)} + \left(\frac{\text{Var(Test)}}{\text{No. of Tests}} \right)} \right) \\ &= 100 + (130 - 100) \times \left(\frac{225}{225 + \left(\frac{25}{1} \right)} \right) = 127 \end{aligned}$$

The best prediction of this individual's true IQ is not 130 but rather 127. Why is this so? This expression for the conditional mean of true IQ given IQ test score may be obtained from the joint distribution of true IQ and IQ test score if both true IQ and the errors of the test are assumed to be normally distributed (Searle, 1971; p. 461). Note that this prediction of true IQ is pulled ever closer to the population mean as the ratio of test error variance to population variance increases or as N becomes smaller. Thus, if a little information is available, a prediction close to the population mean tends to be best. If more information is available, a prediction closer to the sample mean is best. This pulling of the prediction closer to the mean as a function of distance, ratio of the variances, and quantity of information is the essence of the BLUP concept. The concept of BLUP offers an explanation of and a solution to the "regression to the mean" problem.

The Problem of Missing Data. In the original Knox County study, the gain in scale score points for each student was calculated for each student and was used as the response variable in the mixed model equations. This rather simplistic approach

was sufficient to establish the feasibility of the methodology. However, to calculate the gain for each student over multiple years requires no missing data for all year-academic subject combinations. This requirement insures the undesirable result that only a small fraction of student outcomes will be included in an assessment process.

In later work using mixed model methodology, it has been found that complete information for each student is not necessary to provide estimates of the influence of teachers on the gain of a population of students.

Consider the following model to be applied to the data from one specific school system:

$$Y(ijkl) = \mu(ij) + \text{year*subject*teacher}(ijk) + e(ijkl)$$

where

- $Y(ijkl)$ = the student record for the i th year, the j th academic subject, the k th teacher and the l th student record within year, subject, teacher.
- $\mu(ij)$ = the population mean within the i th year and the j th subject.
- $\text{year*subject*teacher}(ijk)$ = the k th teacher in the i th year and the j th subject.
- $e(ijkl)$ = deviation of the l th student score around the year*subject*teacher(ijk) subgroup mean.

Now, consider $\mu(ij)$ to be fixed, and the $\text{year*subject*teacher}(ijk)$ to be random with $\mathbf{G}$ as the variance-covariance matrix among the year*subject*teacher combinations. Let $\mathbf{R}$ be the variance covariance matrix among student records within year*subject*teacher.

The fixed portion of the solution to these mixed model equations will contain the estimated means for the year*subject combinations. The random portion of the solution to these equations will contain BLUP for each year*subject*teacher combination. Directly from this part of the solution vector is available a profile among teachers for each subject each year. These numbers reflect relative differences but are not at this point interpretable as gains. However, these numbers can be scaled to directly reflect gains. The third property (see the section "General Form of Henderson's Mixed Model Equations (MME)," above) of a solution to the mixed model equations is (3) $\mathbf{K}'\mathbf{b} + \mathbf{M}'\mathbf{u}$ is BLUP of $\mathbf{K}'\mathbf{B} + \mathbf{M}'\mathbf{U}$ provided $\mathbf{K}'\mathbf{B}$ is estimable. Thus, by choosing $\mathbf{K}$ and $\mathbf{M}$ appropriately, then BLUP for each teacher's gain is available with its standard error (property 2c).

Property (3) of the solution offers another powerful advantage. By choosing $\mathbf{K}$ and $\mathbf{M}$, teachers can be profiled as math teachers, as reading teachers, as language teachers, etc. or all subjects can be combined to form an overall profile merely by changing $\mathbf{K}$ and $\mathbf{M}$.

Another powerful advantage to this approach is that many different modes of classroom instruction can be accommodated by assigning the teacher of record to each student record within the Z-matrix. It does not matter if a child is in a self-contained classroom, a departmentalized school, or in a team-teaching situation. If the Z-matrix is encoded properly, then BLUP is provided for each teacher. Also, if teachers have assignments over grades each year (i.e., one section of fourth grade math and three sections of fifth grade math), then all information contributes to BLUP. This is also true for teachers changing assignments over years.

Summary

The Tennessee Value-Added Assessment System circumvents many of the problems associated with the use of student achievement data in assessment of school systems, schools, and teachers by relying on the scale scores that indicate gains students make from year to year, regardless of the point at which the student enters the classroom. Three previous studies indicate that the influence of teachers and schools on the rate of student gain were independent of the confounding of socioeconomic factors. The reports of Tennessee school system effects released April 1, 1993, confirm the earlier findings in that the school system cumulative gains for each of the five subjects were uncorrelated with the percent of students receiving free and reduced-cost meals within the system. Also, the cumulative gains for all subjects were found to be uncorrelated with the racial composition of the student body within school systems. Even so, it may be found that some socioeconomic confoundings could surface in the future that would necessitate the inclusion of appropriate covariables in the mixed model equations. Current findings suggest that the number of needed covariables will be relatively small, if any; however, TVAAS readily accommodates such inclusion.

The mixed model methodology upon which TVAAS relies addresses major problems in using student achievement data in educational assessment. Among these are missing student data, diversity of teaching modes, and the regression to the mean problem. The regression to the mean question is dealt with using the concept of Best Linear Unbiased Prediction. The problems of missing student data and diversity of teaching modes are alleviated by retaining the five most current years of data for students and teachers to be included in the mixed model process (Sanders, 1989). By using all of this information for each child and by fitting all the data from teachers over subjects and grades simultaneously, considerable robustness is achieved. This robustness has been confirmed using computer simulations to evaluate "worst case scenarios."

To fit these models to the student data for each school system within a state necessitates monumental computing efforts. For TVAAS to accomplish this task,

it has been necessary to develop a software system to contend with the simultaneous computation of tens of thousands of equations. Each year, as new data are added to the system, solutions to the mixed model equations are newly obtained. Dr. Arnold M. Saxton, Dr. Boyd L. Dearden, Mr. John F. Schneider, and Mr. S. Paul Wright have worked as a team to develop the software and hardware configurations to complete the computations. This team has also developed the reports that were distributed to Tennessee's school systems and has begun analysis of gain patterns that have emerged from the data.

Even though the first reports were issued only a few weeks previous to this writing, many educators had already acknowledged the diagnostic value of the data they have received. It is perhaps here that the impact of TVAAS will be felt most fully. The vast database is yielding far more than assessment data. Because it encompasses so much student data, educational findings that were invisible in the past are now readily apparent. For instance, it was noted that there was a "dip" in scores in the sixth and seventh grades across the state. When the data for homogeneous systems—those systems where all students changed schools in the sixth grade and those systems where all students changed grades in the seventh grade—were aggregated, it was found that gain scores dropped dramatically the year following the school change. Future analysis of school change in systems where a variety of configurations exists is forthcoming. If the pattern persists, it will then be necessary to determine why a change of schools is associated with a drop in gain.

Many other patterns are emerging that bear investigation. Future areas of exploration may include the effects of teaching mode: cooperative learning, whole language, team teaching, etc.; class size; textbook adoptions; funding; technology; curricular innovations and many other factors.

TVAAS offers insight and perspective in the pursuit of educational improvement. It provides a solid basis from which change can be rationally undertaken. The academic gains our students make is the measure of our success as well as theirs.

References

- Astin, A. W. (1982). *Excellence and equity in American education*. Washington, DC: National Commission on Excellence in Education. (ERIC Document Reproduction Service No. ED 227 098).
- Bingham, R. D., Heywood, J. S., & White, S. B. (1991). Evaluating schools and teachers based on student performance. *Evaluation Review*, 15, 191-218.
- Harville, D. A. (1976). Extension of the Gauss-Markov theorem to include the estimation of random effects. *The Annals of Statistics*, 4, 384-395.

- Henderson, C. R. (1973). Sire evaluation and genetic trends. *Proceedings of the Animal Breeding and Genetic Symposium in Honor of Dr. Jay L. Lush*. Champaign, IL: ASAS and ADSA, p. 10-41.
- _____. (1984). *Applications of linear models in animal breeding*. Guelph, Canada: University of Guelph.
- McLean, R. A., & Sanders, W. L. (1984). *Objective component of teacher evaluation—A feasibility study* (Working Paper No. 199). Knoxville, TN: University of Tennessee, College of Business Administration.
- McLean, R. A., Sanders, W. L., & Stroup, W. W. (1991). A unified approach to mixed linear models. *The American Statistician*, 45, 54-64.
- Millman, J. (Ed). (1981). *Handbook of teacher evaluation*. Beverly Hills, CA: Sage.
- Mood, A. M. (1950). *Introduction to the theory of statistics*. New York: McGraw-Hill.
- Sanders, W. L. (1989). A multivariate mixed model. In *Applications of mixed models in agriculture and related disciplines*. Southern Cooperative Series Bulletin No. 343. Baton Rouge, LA: Louisiana Agricultural Experiment Station, pp. 138-144.
- Searle, S. R. (1971). *Linear models*. New York: John Wiley.
- Tennessee State Board of Education. (1992). *The master plan for Tennessee schools 1993*. Nashville, TN: Author.
- Williams, P. L. (1989). *Using customized standardized tests* (Contract No. R-88-062003). Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education. (ERIC Digest No. ED 314429).

An Accountability System Featuring Both "Value-Added" and Product Measures of Schooling

By William Webster and Robert Mendro, Dallas Independent School District

As the nation progresses through the decade of the nineties, there is increased pressure from many segments of society for better educational accountability. This desire for accountability is often accompanied by societal skepticism of educators and the quality of the job that they are perceived to be doing. This perception has been fueled in recent years by the White House and the Department of Education and often used to support an educational agenda that pushes vouchers. The Texas Education Agency, as well as the education agencies of at least 40 other states, has initiated programs that have increased focus on educational outcomes (Duttweiler

& Ramos, 1966; Southern Regional Education Board, 1990). At the national level there is serious talk of a national achievement test (AMERICA 2000, 1991). In Dallas, a group of citizens appointed by the Board of Education developed a comprehensive plan for the improvement of Dallas Schools (Commission for Educational Excellence, 1991). This plan called for rapid conversion from a school system to a system of schools and highlighted accountability as the linchpin for improvement.

The accountability system that is being implemented in the Dallas Independent School District (DISD) and is the subject of this paper is a three tier system. The first tier focuses at the school level. Under the District's plan to move from a school system to a system of schools over a five-year period, each school is held responsible and accountable for many aspects of its own operation. School Improvement Plans (SIP) are the vehicles through which this is accomplished. The second tier of the system involves the District Improvement Plan (DIP). The DIP sets the desired levels on District accountability objectives and specifies how Central Office Divisions support the schools. The third tier involves school effectiveness indices. These indices take into consideration important student background variables and provide information on how effective schools are with the students they serve. The SIP and DIP components of the system focus on the end product of schooling, while the indices provide a "value-added" component to the system.

One of the major concerns related to accountability systems is that of fairness. Educators who are caught up in the accountability movement have a right to know that the standards by which they are judged are fair. The system outlined in this paper attempts to incorporate fairness as defined by the *Standards for Evaluations of Educational Programs, Projects, and Materials* (Joint Committee on Standards for Educational Evaluation, 1981) and the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1985). Where appropriate, this system will be compared to the accountability system being promulgated by the state of Texas and relative strengths and weaknesses enumerated.⁵

School-Centered Education

Accountability Indicators. The District's School-Centered Education Plan focuses control of most available resources and all instructional decisions at the local school level (Edwards, 1991). The only decisions that school level committees are not empowered to make are those involving the nature and magnitude of outcomes

⁵ The Texas Education Agency's (TEA) accountability system is similar in many respects to other state systems. The TEA system is used for comparison purposes because the authors are very familiar with it.

for which they are being held accountable. An extremely important step in the school improvement process is the determination of important performance indicators that will inform educators, parents, and community members whether or not students are making satisfactory progress in the key developmental pathways that they believe are critical for academic learning. These performance indicators are determined by an Accountability Task Force and influenced by the state's Academic Excellence Indicator System. The Academic Excellence Indicator System is the basis for school accreditation in the state of Texas. The accountability indicators are consistent across the three tiers of the accountability system.

The Accountability Task Force. The Accountability Task Force is a 27 member committee, appointed by the Board of Education, charged with the responsibility of overseeing the District's accountability system. The membership includes 4 elementary teachers, 3 middle school teachers, 4 high school teachers, 4 principals, 4 parents, 5 members of the business community, and 3 central office administrators. In addition, the various employee organizations each have an ex officio member on the task force. This task force deals with many aspects of the accountability system including methodology, testing, determining and weighting important performance variables, and determining the rules for financial awards that are related to the accountability system. The Accountability Task Force also hears any concerns or grievances relative to the accountability system.

The Comer Model. The DISD is implementing School-Centered Education through the Yale Child Study Center School Development Program (Comer, 1988). Under this model, the principal, parents, and staff are involved in school decision making and governance through a School-Community Council (SCC), which makes all relevant decisions about school operations. A number of committees can exist at each school, but the SCC and its committees must take responsibility for curriculum, instruction, assessment other than systemwide accountability measures, parent and staff skills development, school-community socialization and interaction, public relations, evaluation, and modification. At the high school level these committees include students.

Regardless of the structure, the evaluation functions that are undertaken at the school level include the development of a SIP; the interpretation of formative data for use in problem-solving and of summative data for use in refocusing priorities, programs, and resources; the development of an implementation record of the various projects and programs within the school, including monitoring the implementation of the SIP; and the coordination of all school-based action research. Central Office research staff provide school personnel with training regarding how to accomplish many of the aforementioned tasks.

The School Improvement Process

Figure 4-4 provides a schematic depicting how the school improvement process functions within the parameters of site-based decision making. Each school receives an annual needs assessment specifying school levels on important outcome variables. The important outcomes of instruction are determined through districtwide assessments of all of the groups involved in the educational process. School program planning is implemented at the school level by the School-Community Council. Planning focuses on determining the best method to proceed from current levels of important outcomes to desired levels of those outcomes and culminates in the production of a strategic plan, the SIP.

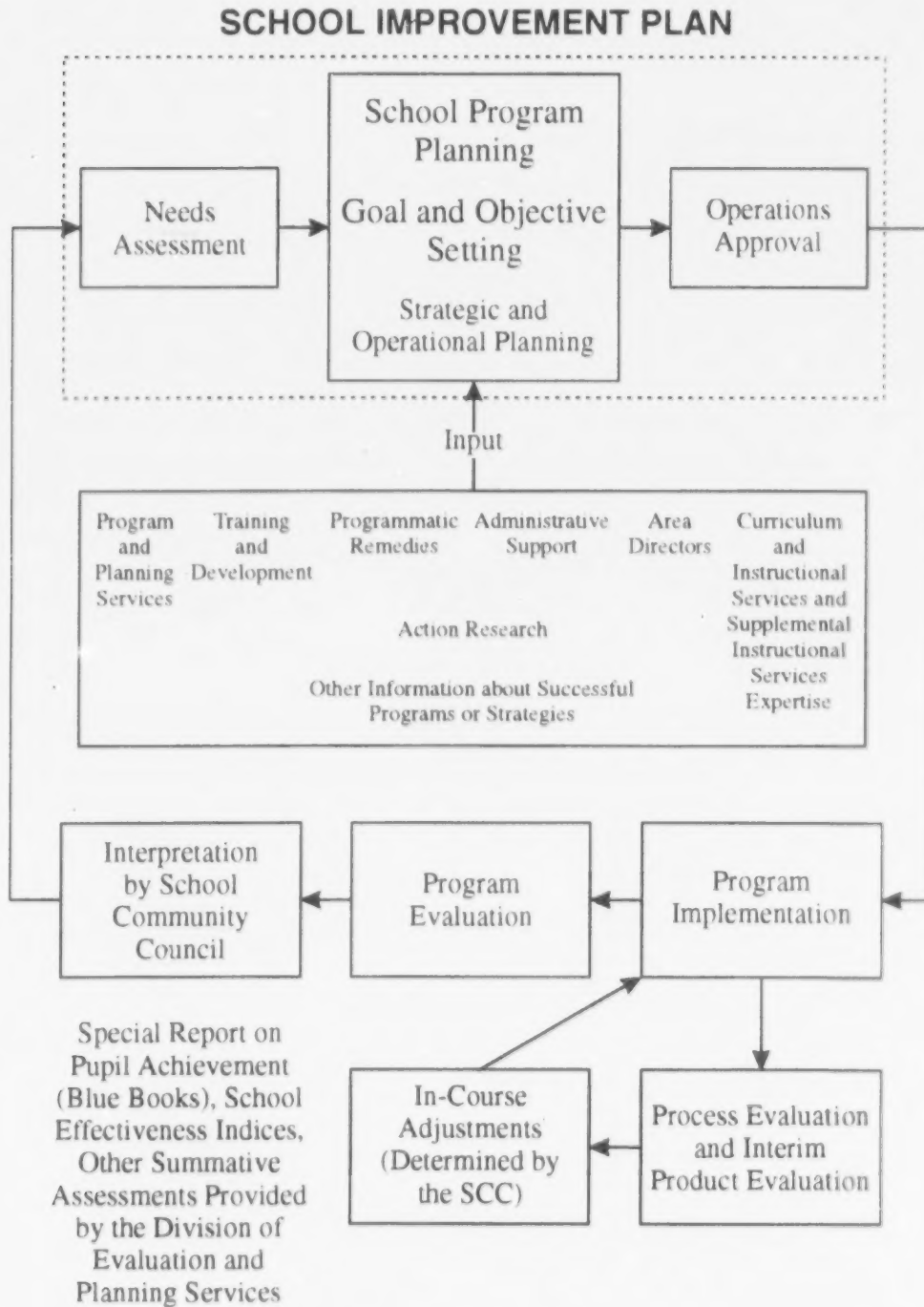
Specifically, once the needs assessment has identified needs, school staff must prioritize those needs and focus on reducing the discrepancy between desired and existing outcomes by establishing goals for those needs that receive highest priority. Once priorities are established, schools must determine methods of resource utilization for accomplishing program goals.

School-centered education does not assume that local building staffs necessarily know how to solve all of their problems. It does, however, place decision-making responsibility and accountability at the local level. Central staff become resources to the schools whose function it is to provide viable alternatives to solving school problems. The principal is ultimately responsible and accountable for meeting the important objectives of instruction. Central staff are responsible and accountable for providing viable alternatives for consideration by school staff and the School-Community Council. This procedure is the input evaluation phase of the school improvement process and will only work if Central Office divisions are competent and can supply the needed expertise. If the needed expertise does not reside in the appropriate Central Office divisions, schools will not request needed services and the entire system will probably fail.

After the collection of relevant input information feeding a preliminary program-planning stage, the School-Community Council determines whether or not sufficient resources are available to make the desired changes. Quite often, sufficient resources are not available and some compromise is necessary. In many cases, the lack of resources is not limited to the realm of cost and political feasibility, but rather stems from an insufficient base of knowledge. Thus, educators are often in the position of having sufficient material resources but insufficient information resources. Once these decisions are made, the School Improvement Plan is complete. Operational review and approval is necessary to insure that the school stays within its resources and that adherence to federal, state, and local policies is maintained.

The program implementation phase is then entered and the individual school staff is responsible for providing continuous formative feedback relative to program

Figure 4-4. Schematic Depicting the School Improvement Process



implementation. This feedback falls primarily into two categories—process evaluation and interim product evaluation. Process evaluation has three major objectives: (1) the detection or prediction of defects in procedural design or its implementation during program implementation stages, (2) the provision of information for programmed decisions, and (3) the maintenance of a record of the implementation procedure as it occurs (Stufflebeam et al., 1971). Thus, process evaluation information keeps the School-Community Council informed of the extent to which program implementation conforms to specifications and, from an evaluation standpoint, guards against the evaluation of a fictitious event. It also provides a record of implementation that can be cross-indexed to program effect.

Much of the process evaluation that was at one time implemented by DISD evaluation personnel now must be implemented at the local level. This is consistent with the accountability emphasis that is currently the philosophy of District management and the community. Since process evaluation is extremely expensive, many of the cutbacks in research and evaluation activities over the past few years have been in the area of process evaluation.

Interim product evaluation provides periodic feedback to the School-Community Council relative to the attainment of specific subobjectives during the implementation phase. Thus, process and interim product evaluation reports inform program management as to implementation and goal-attainment levels while program adjustments are still feasible. Much of the interim product evaluation can be done through portfolios of student work, performance testing, protocol analysis, and teacher-made tests, measures that are not available through systemwide data. Teacher evaluation consultants from each school are trained in these techniques.⁶ In cases where serious needs are identified by interim product evaluation reports, tactical plans are developed as supplements to the SIP to meet these needs.

Local school staffs are also encouraged and trained to design, implement, and interpret action research studies. With the movement of the District to site-based management and the related reduction of Central Office staff, it is impossible to supply school staffs with centrally produced information pertaining to their many and varied needs. Action research is a process for problem solving that is designed and implemented at the local building level. It is a process of taking and studying

⁶ Evaluation consultants are teachers who are trained by the Division of Evaluation and Planning Services to provide evaluation and data interpretation services at the school level. Throughout the school year, the consultants participate in performance-based assessments so that they may learn to apply formative evaluation techniques to their campus' school improvement plans. They identify areas for school improvement, describe program activities, and periodically report information on program impact. This performance-based assessment thoroughly prepares consultants to design defensible evaluations, to measure program implementation, to identify appropriate instrumentation, to assess program impact, and to compile and present reports for school improvement.

action and its corresponding consequences so that more effective action may be taken (Lewin, 1946; Town, 1973). Expressed sequentially, action research requires a continuous recycling through four steps: (1) identification of needs, (2) development of plans of action to address these needs, (3) execution of these plans of action, and (4) formative evaluation of these plans. In open organizations such as schools, the strength of action research lies in its implementation by the organizations' members at their respective work sites. In effect, members of the organization actively learn while they study problems in contexts that they generally perceive as relevant and important. The results are used to supplement the more formal information available from the District's evaluation department.

Upon completion of a given cycle of program implementation, usually one year, a series of summative product evaluation reports are prepared. These reports take the form of the "Special Report on Pupil Achievement (REIS91-102)," a school-level report that provides up to four years of disaggregated data on all relevant outcome and input variables and is used to determine whether or not schools met their SIP goals, School Effectiveness Indices, and program evaluation reports disaggregated by school. These reports, as well as relevant action research studies compiled by school staff, become the needs assessments for the next year's program adjustments.

The District Improvement Plan

The District Improvement Plan (DIP) presents targets and corresponding strategic plans of action with a multiyear planning horizon. Since the District has a number of concerned audiences, the plan meets the accountability objectives and strategic planning requirements of the General Superintendent, the Board of Education, the Texas Education Agency, and the United States District Court. The DIP meets the four major requirements of a strategic planning system in that it receives input from all District departments and campuses, it sets accountability targets and minimum standards of performance for the District and each of its schools, it provides systemwide plans of action for meeting the major targets of the District, and it specifies the methodology required for monitoring its implementation.

The DIP contains the strategic plans of each of the District's support divisions relative to their contributions to meeting each of the District's targets. It also contains the desired levels of District outcomes in the final target year and the intermediate steps necessary to get from baseline levels to those desired outcomes. It is directly related to the SIPs in that outcome levels that are specified in each SIP are those levels that will help the District reach its goals. The DIP sets the criterion level for desired outcomes. Goals are absolute. All schools could make them or no

Table 4-2. DIP Target Areas

<i>Goal #</i>	<i>Goal</i>
1.	Improve language arts skills (vocabulary, reading, oral competency, and writing skills)
2.	Improve mathematics problem solving, concept, and computation skills
3.	Increase parent/community involvement
4.	Improve school climate and safety.
5.	Improve attendance (student and teacher)
6.	Meet accreditation requirements/address citations
7.	Increase promotion/course passing rate.
8.	Increase secondary school enrollment in advanced courses, diploma plans, and honors programs
9.	Increase high school graduation rate
10.	Increase college entrance test participation/performance

schools could make them; that is, target accomplishment is not determined by a norm group.

Targets. The DIP is organized around one systemwide enabling target, staff development, and ten outcome targets that focus directly on the District's priorities. Table 4-2 shows the areas in which District targets currently exist.

DIP Content. Each of the ten targeted outcomes has a strategic plan of action for meeting each target. The plans of action include the following elements:

- *Need* — a needs assessment summary describing the current status of the target.
- *Goal* — reference to the District's minimum accountability and accreditation objectives or other standard of performance that will be met by implementing the plan.
- *Narrative of Strategy* — a summary of what will be done to address the target.
- *Waiver* - a specification of waivers required to implement the strategy.
- *Activities/Time Lines/Divisions Responsible* — activities, corresponding time lines, and divisions responsible for meeting the District's targets.

Table 4-3. Formative and Summative Indicators Available to DISD Schools

<i>Indicators</i>	<i>Goal(s) Impacted</i>	<i>Date Available</i>
*TAAS results disaggregated by demographic variables** Demographic variables include gender, ethnicity, free or reduced lunch and LEP (E)	1, 2, 6	Spring (grades 3-8, 10)
ITBS/TAP results disaggregated by demographic variables (E)**	1, 2, 6	Spring (grades 1-9)
*ACP results disaggregated by teacher and skills (E)**	1, 2, 6, 7	Fall & spring (grades 9-12)
Reconstituted TAAS, ITBS, TAP, data (class lists and skills analyses) (E)	1, 2, 7	End of fourth week of school
Disaggregated test data by program (chapter 1, reading improvement, bilingual, etc.) by school (E)	1, 2, 6	Fall
Portfolios of student work (C)	1, 2	Local option
Performance testing (C,E)	1, 2	Local option
Protocol analysis (C)	1, 2	Local option
Teacher-made tests (C)	1, 2	Local determination
Teacher satisfaction with teaching, ranking of importance of educational goals, perception of teacher influence, and degree of seriousness of school-wide issues (E)	1, 2, 4, 5, 7, 9	Winter (all grades)
Student to volunteer ratio (E,C)	3	Fall
Volunteer hours-to-students (E,C)	3	Fall
Parental involvement log (C)	3	Local option
Parent school expectations, perception of school climate, needs, involvement/participation (E)	3, 4	Winter (all grades)
*Student and teacher attendance (e,c)	1, 2, 5	Each six-week period
Teacher grade distributions (E,C)	1, 2, 6, 7, 9	Each six-week period
School effectiveness indices (E)	1, 2, 5, 9, 10	September
School effectiveness indices disaggregated by student group (E)	1, 2, 5, 9, 10	September

Table 4-3. (continued)

<i>Indicators</i>	<i>Goal(s) Impacted</i>	<i>Date Available</i>
Student satisfaction with learning, academic self-concept, family emphasis on education, cohesion	1, 2, 4	Winter (grades 4-12)
Teacher climate survey (E) (8 scales)	4	Provided on request
Student Climate Survey, Grades 4-12 (E)	4	Provided on request
Principal perceptions of effectiveness of training services, time on task, school-wide issues, decentralization (E)	4	Winter (all grades)
Sociogram of Informal Interaction (lunch, recess, faculty meetings, etc.) (C)	4	Local option
School-Community Council Survey (E)	4	Fall and spring
Assistance and Consultation Team (ACT) Surveys (global issues, case management, training on mental health principles) (E)	4	Fall and spring
Measures of Mobility and Stability (E)	5	Fall
Percent Eligible Tested versus Average Daily Attendance (E)	5	Fall
Monitoring of Local School Accreditation Remedies (C)	6	Fall
Monitoring of Implementation of Local School Programs (C)	7	Local option
Monitoring of Instructional Delivery (C)	1, 2, 4, 6, 7	Local determination
Student Retention Rate (E)	7	Fall
*Student enrollment in advanced placement and honors courses (E,C)	8	Fall, spring
*Student enrollment in advanced diploma plans (E,C)	8	Fall, spring
Survey of student course interest (grades 7-12) (E)	8, 9	Provided on request
*Dropout rate (E)	9	December
Graduation rate (E)	9	Fall
*SAT/ACT participation rates (E,C)	10	Fall

Table 4-3. (continued)

<i>Indicators</i>	<i>Goal(s) Impacted</i>	<i>Date Available</i>
*SAT/ACT scores (E)	10	Fall
Graduate follow-up (E)	9	Fall
Student post-graduate pursuits (E)	8, 9, 10	Fall
PSAT participation rates (E)	10	Fall
PSAT scores (E)	10	Fall

* An Academic Excellence Indicator in the State Accreditation System

** TAAS is the *Texas Assessment of Academic Skills*, a State-administered criterion-referenced test. ITBS is the *Iowa Tests of Basic Skills*. TAP is the *Tests of Achievement and Proficiency*. ACPs are 143 criterion-referenced course exams, grades 9-12.

- *Monitoring* — the methodology for directing, assessing, adjusting, and documenting formative activities to meet the goal.
- *Resource Implications* — a summary of the distribution (e.g., monies, personnel) changes required to implement the strategies.

One problem that plagues these types of absolute systems is the problem of setting meaningful goals. The issue of low expectations versus lofty goals that are unattainable comes into play. This is why there must be other components of the system to make it fair and useful. The effectiveness index component of the system, discussed in the next section of the paper, can be used to establish meaningful absolute goals by basing those goals on best practice the previous year. That is, those schools that rank high on the effectiveness indices can be used to demonstrate achievable goals for other schools.

Authentic Assessment and Performance Testing. Schools are encouraged to use portfolios, protocol analysis, and other forms of authentic assessment in monitoring their programs. This information can then be used to provide evidence of accomplishment in instances where the more standard types of assessment fail to show progress. Performance testing was at one time being built into the District's *Assessment of Course Performance (ACP)* test. The ACPs are final standard examinations in 72 courses, grades 9-12. One hour was to be multiple choice while the other hour was to be performance tests. These tests were developed by the evaluation division and had detailed scoring protocols. The performance portion of the tests would have been scored by teachers, with random scoring being done by the evaluation department. Performance testing was subsequently eliminated by District administration as being too time consuming.

Table 4-4. An Example of a High School Accountability and Accreditation Profile*

<i>Outcome Variables</i>	<i>Baseline</i>	<i>Year 1</i>	<i>Year 2</i>	<i>Year 3</i>	<i>Year 4</i>
GRADUATION RATE (5 yr. %) (ACCRED. GRDTN RT)	52	54	57	59	61
% ADA—STUDENTS (CLI. ATTNDNC, ACCRED)	90.3	91.3	92.1	92.9	93.6
% ADA—TEACHERS (CLIMATE, ATTENDANCE)	96.5	97.5	98.0	98.0	98.0
% FROSH ADV GRAD PLANS (ACCRED. ENR ADV PLANS)	37	40	43	46	49
% SENIORS TAKING SAT/ACT (ACCRED. COL TSTS)	57	59	61	63	65
% SR > 700 SAT (16 ACT) (R, W, M, ACCRED, COL TSTS)	72	73	75	76	77
% SR > 700 SAT (21 ACT) (R, W, M, ACCRED, COL TSTS)	27	31	34	37	41
% SR > 1300 SAT (27 ACT) (R, W, M, ACCRED, COL TSTS)	4	9	13	18	22
% GRADUATES CONT EDUC (PRMTN/GRDTN RT, COL TSTS)	48	51	53	55	58
% SRV LEP ≥ 40 R&L POST (R, W)	5	10	14	19	23
% PASSING ALL COURSES (PRMTN RT)	53	55	58	60	62
% IN HONORS/AP/PRE-HNRS (ACCRED, ENR ADV PLANS)	22	26	30	33	36
DROPOUT RATE (%) (COMM, CLI, ACCRED, GRDTN RT)	7.0	6.5	6.0	5.5	5.0
% PASSING TAAS, Grade 10	20	28	35	42	48
% PASSING CURRICULUM REFERENCED TEST					
Biology	60.5	62.5	64.4	66.4	68.3
Chemistry I	65.0	66.5	68.0	69.5	71.0
Physics	70.9	71.9	72.9	73.9	74.9
U.S. History	63.6	65.2	66.9	68.5	70.2
Economics	68.1	69.3	70.5	71.7	72.9
Trigonometry	72.4	73.4	74.4	75.4	76.4
English I	69.1	73.4	74.4	75.4	76.4
NORM REFERENCED TESTS** TAP ≥ 25					
Grade 9 Reading (TAP Median)	82	75	75	75	75
Grade 9 Reading (TAP ≥ 75)	62	63	64	65	65
Grade 9 Reading	33	25	25	25	25

* The profiles include many more variables. This figure is for illustrative purposes.

** Because the state changes the norm-referenced test every year, District goals in this area are to mirror the national norm group.

While it is not certain that the necessary reliability across scorers and tasks on the performance tests would have been attainable, it is important that the message be communicated to teachers that the kinds of skills and activities measured by performance tests are the kinds of skills and activities that the District wants them to teach their students. Thus, performance testing is more of a curriculum issue than an assessment issue. Early evidence on performance tests suggests that they are much more difficult than the average multiple choice tests (Dryden, 1991). Table 4-3 shows the formative and summative data currently available to the schools. Indicators that are collected centrally and provided to schools are specified with an "E." Formative indicators that should be part of a school's "action research" process are specified with a "C." State academic excellence indicators are asterisked, while variables that are or will be outcome variables in the effectiveness indices are marked with a #.

Obviously, a great deal of training must occur if school staffs are to utilize available data and objectively collect and interpret additional data for aid in improving their schools. Training modules for school staffs are currently being developed in keeping and scoring student portfolios of work, designing and scoring performance tests, conducting protocol analysis, developing teacher-made tests, interpreting and using data, and designing and conducting action research.

Accountability without information for diagnosis and improvement is of limited utility. In designing an accountability system, it is important to analyze data needs at each point in the organization. Data needs at the teacher level should be identified and those data aggregated upward and summarized to meet information demands at each successive level of the organization. It is essential that the system provide teachers with the information necessary to improve instruction. Without instructional improvement, accountability alone cannot improve a school system.

Table 4-4 shows an example of the operationalization of the DIP targets. Each school receives its own data on each of these targets and is responsible for achieving its targeted outcomes. The targets are criterion-referenced in the sense that the schools have absolute goals and can concentrate resources on attempting to achieve those goals. One major problem with these goals is that they are not empirically established.

School Effectiveness Indices

The final tier of the accountability system is the most important from the standpoint of defining and rewarding outstanding schools. Inherent in the task of identifying outstanding schools are two complex issues:

- how to define effectiveness
- how to develop a model to assess effectiveness

In an attempt to provide a better definition of effectiveness and respond to the narrowly focused concern of earlier effective schools research, Murnane (1987), David (1987), and others have been proponents for developing an expanded number of outcome indicators. In addition, Oakes (1989), David (1987), and Cohen (1986) have argued the importance of incorporating input and process/context indicators as important aspects of better accountability mechanisms.

Possible input indicators often include school enrollment, socioeconomic/ethnic composition, proportion of limited-English-speaking children, enrollments in categorical programs, staff characteristics, and financial resources. Process indicators describe what is being taught, the way it is being taught, and include consensus on school goals, instructional leadership, opportunity to learn, school climate, staff development, and collegial interaction among teachers. Outcome indicators are usually related to capturing the results of school on students or providing information about other definitions of "good schooling," and may include student academic performance, teacher and student attendance rates, dropout and completion rates, performance of students at the next level of schooling, parent and student satisfaction, percent completing advanced courses, college attendance, and individual school goals (David, 1987; Oakes, 1989; Olson & Webster, 1990; Pollard, 1987; Shavelson, McDonnell, Oakes, & Carey, 1987).

The Academic Excellence Indicator System. The Texas Education Agency (TEA), like many other state education agencies, has its own accountability system. This system is called the Academic Excellence Indicator System and includes the variables that are asterisked in Figure 3. It reports data on districts in both a cross-sectional and cross-sectionally longitudinal manner and purportedly allows for the comparison of districts to "like" districts and to the state as a whole. This system has many flaws.⁷

If one overlooks the flaws in basic measurement that are often present in state testing programs, flaws that extend all the way from unreliable tests to tests that are not scaled yet used to make quasi-longitudinal comparisons, the technique of comparing schools based on unadjusted outcome measures usually adversely affects schools with population demographics that differ from the norm. This fact was graphically illustrated relative to ethnic background and SAT scores in a recent article by Richard Jaeger (1992). The nonstatistical technique of comparing schools with similar characteristics is one solution for cases involving a limited number of grouping characteristics; however, this approach has serious limitations when there is consistent one-directional variance on the grouping characteristics within group.

⁷ The TEA has subsequently dropped the comparison of districts to "like" districts from its accountability system. This system has been replaced with one encompassing "world class standards."

To illustrate this point, examine the group wherein the DISD was classified in the Academic Excellence Indicator Report published by the Texas Education Agency in 1992. The DISD was 15.9 percent white, the comparison group was 20.9 percent white. The DISD was 45.5 percent African American, the comparison group was 38.8 percent African American. The DISD was 66.5 percent poor, the comparison group was 55.5 percent poor. The DISD was 19.3 percent LEP, the comparison group was 17.4 percent LEP. Thus, on every important variable, the DISD had the group that performed most poorly on the TAAS statewide and, not surprisingly, performed lower than the comparison group. Yet, when those scores were adjusted for only the ethnic background of students, DISD performed at about state levels (Webster, 1991).

The new TEA accountability and accreditation system still relies on Academic Excellence Indicators but adds a "value-added" component. Unfortunately, this component uses a very crude methodology that makes no attempt to statistically adjust for any student entry variables. While not intending to provide a critique of TEA's accountability system, the system does provide illustrative examples of inappropriate methodology and interpretation.

First, the system provides arbitrary passing criteria with no evidence of predictive or concurrent validity. "Accredited" is set at 25 percent passing each TAAS subtest, "Recognized" at 65 percent, and "Exemplary" at 90 percent. Since there is a published relationship between TAAS scores and ethnic background as well as economic disadvantage, the system guarantees that there will be few predominantly minority economically disadvantaged schools "Recognized" or "Exemplary" regardless of how much they improve their students' achievement levels.

There is a "value-added" component to the system. However, it is not statistically sound and, in fact, requires the schools that are having the most difficulty serving their students to improve the most, regardless of student background. If a school's student population does not pass 25 percent of the items or the scale score equivalent on TAAS, that school's increase in percentage passing must be greater than:

$$(50\%) - \frac{(\% \text{ passing TAAS in 1993})}{5}$$

That is, the decision has been made, without benefit of empirical data or consideration of the difficulty level of each TAAS, that 50 percent passing is the goal. Similar reasoning has taken place at the "Recognized" levels where the criterion is 90 percent instead of 50 percent. What has resulted is an accountability system that is biased against the economically disadvantaged and ethnic minorities and ensures that 100 percent white, noneconomically disadvantaged schools do very well whether or not they are doing anything for their students. When coupled with TEA's refusal to publish relevant statistics on current tests, one is left with an

accountability system that, by careful choice of test items, can at one time present the picture of quality educational improvement in the state, say, for example, in a year of a gubernatorial election, to a system that, in a year when it is politically expedient, supports the movement of funds and students away from the public schools in implementing a mandate for choice. Of course this system would largely support such a mandate in predominantly minority schools.

For accountability purposes, the only fair and equitable method of comparisons among and between schools or districts is one that statistically adjusts measures of important student outcomes by important student inputs that are related to those outcomes but not under the control of the schools. This statistical adjustment may be accomplished in a number of different ways.

Fennessey and Salganik (1983) proposed a model for analyzing instructional program effectiveness within the context of gain scores. The rescaled and adjusted gain score (RAGS) index equalized aggregate net bias from responsiveness to instruction, regression-to-the-mean, and boundary artifacts in all program groups. A crucial assumption to this approach is that any group of students with similar pretests scores will have similar rates of learning and will be subject to the same degree of regression-to-the-mean. While the RAGS procedure is appropriate for program evaluation, it would be difficult to apply in a situation where one is attempting to determine the relative effectiveness of schools with very different student populations.

Another statistical method that has been widely touted as appropriate for incorporating a large number of input and outcome variables in a fair and unbiased manner is multiple regression analysis (Bano, 1985; Felter and Carlson, 1985; Kirst, 1986; Klitgaard & Hall, 1973; MacKenzie, 1983; Saka, 1989). As a simplified illustration, the mean score for an outcome measure such as achievement is predicted after considering such input variables as gender, ethnicity, and socioeconomic level. The equation becomes more accurate if one or more estimates of previous achievement level are included. The difference between predicted and actual achievement, a residual or adjusted score, can then be interpreted as a comparison with other statistically similar schools and as the school's own effect on achievement. It is important to note that a longitudinal database is necessary for these types of studies since cohorts must be used in the analyses. The characteristics of such a data base are detailed in Webster and Schumacher (1973).

There are two obvious approaches that can be used in applying regression analysis to the definition of effective schools. The first involves the disaggregation of all variables to the student level. At the student level only student level characteristics are used and the analysis is done on the individual. Traditional linear model analysis requires four basic assumptions: normality, linearity, homoscedasticity, and independence. Obviously, because students are in the same classes, this approach cannot meet the assumption of independence. However, since we are

working with the entire population, meaning that we are entirely in the domain of descriptive statistics and we are not interested at this point in partitioning teacher effect and school effect, we are not concerned with the assumption of independence or that of normality. The models that are described in the next sections go to great lengths to meet the assumptions of linearity and homoscedasticity. If the linearity assumption is not met, higher order equations are used.

The second alternative would be to aggregate individual-level variables to the higher level and conduct the analysis at the aggregate level. This approach was rejected because it eliminated too much information, produced little variance among schools, and produced much different equations and rankings than did the individual-level analysis (Mendro & Webster, 1993). In addition, the results produced had no face validity.

Another regression-based approach, hierarchical linear modeling (HLM), estimates linear equations that explain outcomes for group members as a function of the characteristics of the group as well as the characteristics of the members. Because HLM involves the prediction of outcomes of members who are nested within groups that in turn may be nested in larger groups, it is well suited for use in education. The nested structure of students within classrooms and classrooms within schools produces a different variance at each level for factors measured at that level. While the current Dallas regression models only use student-level variables in prediction, it is probable that the error terms can be improved through the use of HLM. Bryk et al. (1988) cited four advantages of HLM over regular linear models. First, HLM can explain student outcomes and growth as a function of school-level or classroom-level characteristics while taking into account the variance of student outcomes within schools. Second, it can model the effects of student background variables such as gender, ethnicity, limited-English proficient status, and socioeconomic status on outcomes within schools or classrooms and explain differences in these effects between schools or classrooms using school or classroom characteristics. Third, HLM can model the between- and within-school variance simultaneously and thus produce better estimates of student outcomes. Finally, it can produce better estimates of the predictors of student outcome within classrooms by using information about these relationships gained from other schools and classrooms. HLM models are discussed in the literature under a number of different titles by different authors from a number of diverse disciplines (Bryk & Raudenbush, 1992; Dempster, Rubin, & Tsutakawa, 1981; Elston & Grizzle, 1962; Goldstein, 1987; Henderson, 1984; Laird & Ware, 1982; Longford, 1987; Mason, Wong, & Entwistle, 1984; Rosenberg, 1973). The authors are currently exploring the applicability of HLM (Mendro, Webster, & Bemby, 1994; Webster, Mendro & Ortiz, 1994).

The Anatomy of Effectiveness Indices. The school effectiveness methodology, as implemented in the DISD, defines a school's effectiveness as being associated with exceptional measured performance above or below that which would be expected across the entire District. When a school's population of students departs markedly from its own preestablished trend or from the more general trend of similar students throughout the District, this departure is attributed to school effect. The problem of measuring a school's effect, then, becomes one of establishing the student levels of accomplishment on the various important outcome variables, setting levels of performance based on these expectations, and determining the extent to which its students, on the average, exceed or fall short of expectation. The procedures involve regression analysis to compute prediction equations by grade level for each outcome variable independent of school identification and then using those equations within schools to obtain mean gains over expectations. Relative weights are assigned to the outcomes by the Accountability Task Force. Once weighted levels of performance have been determined, the methodology provides an indicator of how well a school performs relative to other schools throughout the District. To a great extent, the same targets that were used in the SIP and DIP processes were used as outcome variables in the school effectiveness indices. Thus, schools work on improving target variables in an absolute sense through their SIPs and are judged in terms of a normative rank through the effectiveness indices.

School performance on the effectiveness indices is considered in terms of overall District patterns on the important outcome variables. If the District experiences a year of greatly increased achievement, individual school ranks on the effectiveness indices are not so important as long as improvement is shown. The emphasis of the methodology is currently on the valid identification of effective schools, not on explaining their effectiveness through mathematical models such as path analysis or hierarchical linear modeling. Once effective schools are reliably and validly identified, detailed studies can be conducted to attempt to determine process variables that contributed to their effectiveness.

The first step in developing the effectiveness methodology involved what educational practitioners have called "leveling the playing field." The Accountability Task Force was extremely concerned that all schools, regardless of the students they served, had an opportunity to rank high on the effectiveness indices if they improved. Thus, the first step in developing the equations was to eliminate the variance in outcomes accounted for by ethnicity, gender, socioeconomic status, and limited English proficiency status.⁸ To accomplish this each outcome and predictor variable was regressed on the set of background variables and their

⁸ Socioeconomic status was defined by free or reduced lunch, parent education level, household income, and poverty classification.

interactions to produce a set of residuals for each of the predictor and outcome variables. (Webster, Mendro, & Almaguer, 1994).

Once each student's standardized residual values were computed on each of the predictor and criterion variables, each predictor space was divided into 256 arrays and the residuals were standardized. This was done to insure that schools that had unusual numbers of students in certain areas of the predictor space were not ranked based upon differential variance in different arrays. These standardized residuals were then used to develop the next level of equations.

An all possible regressions approach was then used on the residuals of both the outcome and predictor variables. Equations were developed utilizing individual students rather than school means. Where school level variables were used, separate equations were run. Satisfactory prediction was achieved in all cases without having to go back more than one year. This maintained the degrees of freedom associated with the equations. A previous model that was utilized by the District in 1984 used a variant of time-series analysis, but since this model required at least three years of historical data, it suffered from severe subject mortality due to a high student mobility rate (Webster & Olson, 1988).

Again, the predictor space was divided into 256 arrays and standardized for each criterion variable. As before, this was done to insure that schools derived no particular advantage by starting with high-scoring or low-scoring students or with large numbers of students at a particular point in the predictor space. That is, schools were not disadvantaged by differential variance in the predictor space at different points along the regression line.

The individual student residuals were then associated with the schools from which they came. Mean residuals were obtained on each of the criterion variables to provide a gross estimate of school effect. These mean residuals were then multiplied by $\sqrt{n}$ to equalize the variance of the different school means and restandardized to a mean of 50 and a standard deviation of 10. Finally, the mean standardized residuals were multiplied by the weights assigned by the Accountability Task Force and aggregated for each school to produce the final school effectiveness index. Figure 5 shows the 1993-94 weights assigned to various outcome variables by the Accountability Task Force.

Study of Table 4-5 shows that the Texas Assessment of Academic Skills (TAAS) was heavily weighted at all grade levels. This was due to the fact that the Texas Education Agency was using this test in its accreditation system, and failure to master it carried strong sanctions at both the school and individual student level. Other variables included the Iowa Tests of Basic Skills (ITBS), grades 1-9, student promotion rate, grades 1-8, student attendance, grades 1-12; the Spanish Assessment of Basic Education (SABE), grades 1-6; the Assessments of Course Performance (ACP), grades 9-12, graduation rate, grades 9-12; Scholastic Aptitude Test (SAT) percent taking and score, grades 10-12, dropout rate, grades 7-12, percent

Table 4-5. Weighting of Criterion Measures by the Accountability Task Force

Grade	1	2	3	4	5	6	7	8	9	10	11	12
ITBS												
Reading	2	2	2	2	2	2	2	2	4	•	•	•
Math	2	2	2	2	2	2	2	2	4	•	•	•
Promotion Rate	1 per school						1	•	•	•	•	
Attendance	1	1	1	1	1	1	1	1	1	1	1	1
TAAS												
Reading	•	•	5	5	5	5	5	5	•	12	•	•
Writing	•	•	•	5	•	•	•	5	•	12	•	•
Math	•	•	4	4	4	4	4	4	•	12	•	•
Science	•	•	•	1	•	•	•	1	•	•	•	•
Social Studies	•	•	•	1	•	•	•	1	•	•	•	•
SABE	2	2	2	2	2	2	•	•	•	•	•	•
IIACP												
Language Arts	•	•	•	•	•	•	•	•	2	2	2	2
Math	•	•	•	•	•	•	•	•	2	2	2	2
Social Studies	•	•	•	•	•	•	•	•	2	2	2	2
Science	•	•	•	•	•	•	•	•	2	2	2	2
ESOL	•	•	•	•	•	•	•	•	2	2	2	2
Reading	•	•	•	•	•	•	•	•	2	•	•	•
World Language	•	•	•	•	•	•	•	•	2			
Graduation Rate	•	•	•	•	•	•	•	•	5			
SAT % Tested	•	•	•	•	•	•	•	•	•	5		
SAT Score	•	•	•	•	•	•	•	•	•	4		
Dropout Rate	•	•	•	•	•	•	1	1				
Accelerated Courses	•	•	•	•	•	•	5	4				
ACP Honors Course	•	•	•	•	•	•	•	•	3			
Advanced Diploma Plan	•	•	•	•	•	•	•	•	3	2	•	•
PSAT % Tested	•	•	•	•	•	•	•	•	3			
PSAT Score	•	•	•	•	•	•	•	•	2			

enrolled in accelerated courses, grades 7-12; ACP performance in Honors courses, grades 9-12; percent enrolled in Advanced Diploma Plans, grades 9-10; and percent taking and score on the Preliminary Scholastic Aptitude Test (PSAT), grades 9-12.

Dallas Independent School District schools and their staffs were eligible for cash awards for 1993-94 performance based on the school effectiveness methodology under the District's School Performance Improvement Awards Program. In September of 1994, 2.4 million dollars was distributed to effective schools and their employees. Half of the 2.4 million dollars was budgeted by the District, the other half came from the community. To qualify, schools had to exceed prediction on the effectiveness indices, test 95 percent of their eligible students, and outgain the national norm group in at least 50 percent of their cohorts. Once a school was selected as an award winner, the school received \$2,000 for its activity fund, each member of its professional staff received \$1,000, and each member of its support staff received \$500. This program is continuing in 1994-95.

Teacher Effectiveness Indices. Since the teacher is the principal deliverer of instruction to students, it is essential that a method for attributing student outcomes to teachers be developed. Of the numerous methods available for teacher evaluation, student outcome data provide an attractive option for evaluating teacher effectiveness since they are basically objective in nature. One only needs to examine the distributions of teacher evaluations and student achievement in the average urban District to quickly realize that something is wrong with the current system. A number of researchers have enumerated many factors that inhibit the reliable and valid use of student achievement data in an evaluation of teacher performance (Bano, 1985; Dutweiler & Ramos-Cancel, 1986; Grobe, 1992; Haertel, 1986; Koehler, 1985; Murnane & Cohen, 1986; Redfield, 1987).

Some of the most troublesome of these factors include

- Standard measurement instruments are not available for many courses and subject areas. (In Dallas, about 57 percent of high school course sections have ACPs.)
- Reliable performance measures are nonexistent.
- Because of team teaching, pull-out and send-in programs, and other special programs, it is difficult to isolate an individual teacher's effect on individual students. (In Dallas, supplementary teachers account for 30 percent of the teaching force).
- What the student brings to the classroom in terms of ability, home and peer influence, motivation, etc., is very powerful in affecting school outcomes. Attempting to adjust student outcomes based on student inputs at the teacher level creates serious degrees of freedom problems; that is, the resulting estimates are not stable.

The preferred solution to this dilemma is to provide the principal, as the instructional leader and Chief Executive Officer of the school, with a large and diverse set of explanatory variable information to help him or her with the evaluation of teachers. Information includes class improvements on outcome measures over the previous year through the effectiveness index methodology, class item and skills analyses, class background information; and the system allows for teacher-generated information such as protocol analysis. Also provided are standards that communicate progress being made with similar students across the District. The emphasis is currently on diagnosis and improvement, although the information will eventually be used as part of the teacher evaluation system. At that point, Hierarchical Linear Models may be applied to obtain teacher indices. There are currently no plans for rewarding individual teachers based on some type of effectiveness index, since one of the major strengths of the school effectiveness methodology is the staff collegiality that it reinforces. Such collegiality is important in restructuring around Comer's school-centered education model.

Summary

This chapter has described a three-tier accountability system. District goals and desired outcomes are established through a districtwide planning process and operationalized through the District Improvement Plan. Each school's role in helping the District to meet its goals is determined through a School-Community Council that ensures involvement at the local campus level. Accountability is operationalized in a criterion-referenced manner through an analysis of absolute outcomes relative to school and District performance on goals specified in the District Improvement Plan and the School Improvement Plans, and in a norm-referenced manner through school effectiveness indices. Schools and their staffs are eligible for financial awards based on school performance on the effectiveness indices.

Besides providing an objective procedure for identifying effective schools, the program has a number of practical advantages. First, and most important, it is designed to foster teamwork among school staffs within schools. In order to achieve the necessary improvements in student outcomes, school staffs must work together in a coordinated effort. The program does not reward individual competition among teachers within schools.

Second, the program focuses attention on the important outcomes of schooling. The Accountability Task Force, as well as many other groups associated with the schools, are discussing what it is that the schools are about. The process of weighting the outcome variables, a procedure that is done annually, gives many divergent groups the opportunity to share their views relative to the purposes and

importance of schooling. While the accountability system alone will not improve instruction, the curriculum and instructional delivery processes that must be changed to impact the defined outcomes will.

Third, the procedures described afford all schools an opportunity to be distinguished in the awards independently of their student population status on the achievement continuum. The emphasis is on effectiveness with the students who come in the door, not absolute outcome levels. The techniques reward those schools that impact the most students the most positively (Webster, Mendro, & Almaguer, 1993).

Many District and State accountability systems include District and School Improvement Plans that encompass absolute goals. The addition of effectiveness indices make the accountability system valid and fair. Among the advantages of this type of approach are that each school's performance is not judged by simple examination of raw outcome variables, but instead by comparing its student outcome levels with empirically determined expectations based on individual student histories; that schools derive no particular advantages by starting with high-scoring or low scoring students of any particular ethnic or economic group; that schools are only held accountable for the outcome levels of continuously enrolled students, that is, students who have been exposed to their instructional program; that adequate time for test make-up is allowed and schools must test 95 percent of their eligible students; and a Task Force representing all of the important groups that have a stake in schooling determines the important outcomes of schooling and their respective weights in the equations.

References

- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Bano, S. M. (1985). *The logic of teacher incentives*. Washington, DC: National Association of State Boards of Education.
- Berk, R. A. (1984, March). *The use of student achievement test scores as criteria for allocation of teacher merit pay*. National Conference on Merit Pay For Teachers, Sarasota, FL.
- Berk, R. A. (1988). Fifty reasons why student achievement gain does not mean teacher effectiveness. *Journal of Personnel Evaluation in Education*, 1, 345-363.

- Bryk, A. S., Raudenbush, S. W., Seltzer, M., & Congdon, R. (1988). *Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model*. San Diego: Academic Press.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear model: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Cohen, M. (1986). *Designing state education assessment systems*. Study Group on the National Study of Student Achievement.
- Comer, J. P. (1988). Educating poor minority children. *Scientific American*, 259(5), 42-48.
- Commission for Educational Excellence. (1991). Final report. Dallas: DISD.
- David, J. (1987). *Improving education with locally developed indicators*. New Brunswick: Center for Policy Research in Education, Rutgers University.
- Dempster, A. P., Rubin, D. B., & Tsutakawa, R. V. (1981). Estimation in covariance components models. *Journal of the American Statistical Association*, 76, 341-353.
- Dryden, M. (1991). *Evaluation of the 1990-91 south and west Dallas learning centers*. Dallas, TX: Division of Evaluation and Planning Services, DISD.
- Duttweiler, P. C. & Ramos-Cancel, M. L. (1986). *Perspectives on performance-based incentive plans*. Austin, TX: Southwest Educational Laboratory (ERIC ED272511).
- Edwards, M. E. (1991). *School-centered education: A plan for creating a system of schools in the DISD*. Dallas, TX: DISD.
- Edwards, M. E. (1993). *School-centered education: The Dallas model*. Dallas, TX: DISD.
- Elston, R. C., & Grizzle, J. E. (1962). Estimation of time response curves and their confidence bands. *Biometrics*, 18, 148-159.
- Felter, M. (1989). *A method for the construction of differentiated school norms* (ERIC ED312302). American Educational Research Association, San Francisco.
- Felter, M., & Carlson, D. (1985). Identification of exemplary schools on a large scale. In Austin & Gerber (eds.), *Research on exemplary schools*, pp. 83-96. New York: Academic Press.
- Fennessey, J., & Salganik, L. H. (1983). *Credible comparison of instructional program impact: The RAGS procedure*. Educational Measurement Issues and Practice: Johns Hopkins University.
- Goldstein, H. (1987). *Multilevel models in educational and social research*. New York: Oxford University Press.
- Grobe, R. P. (1992). *Using student-achievement data for teacher evaluation*. Dallas, TX: Division of Evaluation and Planning Services, DISD.
- Haertel, E. (1986). The valid use of student performance measures for teacher evaluation. *Education Evaluation and Policy Analysis*, 8, 45-60.

- Haertel, E. (1990). Performance tests, simulations, and other methods. In J. Millman & L. Darling-Hammond (eds.), *The New Handbook of Teacher Evaluation: Assessing Elementary and Secondary School Teachers*, 278-294. Newbury Park, CA: Sage.
- Henderson, C. R. (1984). *Applications of linear models in animal breeding*. Guelph, Canada: University of Guelph.
- Institutional Research Office. (1992). *Special reports on pupil achievement*, REIS92-102, 1-200. Dallas, TX: DISD.
- Jaeger, R. M. (1992). Weak measurement serving presumptive policy. *Kappan*, 74(2), 118-128.
- Joint Committee on Standards for Education Evaluation. (1981). *Standards for evaluations of educational programs, projects, and materials*. New York: McGraw-Hill.
- Kirst, M. (1986). New directions for state education data systems. *Education and Urban Society*, 18(2), 343-357.
- Klitgaard, R. E., & Hall, G. R. (1973). *A statistical search for unusually effective schools*. Santa Monica, CA: Rand Corporation.
- Koehler, V. (1985). National overview: Career ladder and merit-pay plans. In C. Clark (Ed.), *Proceedings of the preventive law institute on career ladder/merit pay*. Austin, TX.: Southwest Educational Development Laboratory (ERIC ED272511).
- Laird, N. M., & Ware, H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38, 963-974.
- Lewin, K. (1946). Action research and minority problems. *Journal of Social Issues*, 2, 58-73.
- Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, 74(4), 817-827.
- MacKenzie, D. (1983). School effectiveness research: A synthesis and assessment. In P. Duttweiler (Ed.), *Educational productivity and school effectiveness*. Austin, Texas: Southwest Educational Development Laboratory.
- Mason, W. M., Wong, G. Y., & Entwistle, B. (1984). Contextual analysis through the multilevel linear model. In Leinhardt (Ed.), *Sociological methodology*, pp. 72-103. San Francisco: Jossey-Bass.
- Mendro, R. L., & Webster, W. J. (1993). *Using school effectiveness indices to identify and reward effective schools*. Rocky Mountain Educational Research Association, Las Cruces, NM.
- Mendro, R. L., Webster, W. J., & Bembry, K. (1994). *An application of hierarchical linear modeling in determining school effectiveness*. American Educational Research Association, San Francisco.

- Murnane, R. J. (1987). Improving education indicators and economic indicators: The same problems? *Educational Evaluation and Policy Analysis*, 9, 101-116.
- Murnane, R. S. (1991). The case for performance-based licensing. *Kappan*, 73(2), 137-142.
- Murnane, R. S., & Cohen, D. K. (1986). Merit pay and the evaluation problem: Why most merit pay plans fail and few survive. *Harvard Educational Review*, 56, 1-17.
- Nicoll, R. (1989). *School accountability in the Mt. Diablo unified school district*. American Educational Research Association, San Francisco.
- Oakes, J. (1989). What education indicators? The case for assessing the school context. *Educational Evaluation and Policy Analysis*, 11, 181-199.
- Olson, G. H., & Webster, W. J. (1986). *Measuring school effectiveness: A three-year study*. American Educational Research Association, San Francisco.
- Olson, G. H., & Webster, W. J. (1990). *Proposed procedures for measuring school effects in the Dallas Independent School District*. DISD, REIS 90-140.
- Pollard, J. (1987). *Viewpoints from selected states on accreditation and accountability*. Austin, TX: Southwest Educational Development Laboratory.
- Redfield, D. (1987). *Expected student achievement as a potential factor for assessing teacher effectiveness*. ERIC ED 290764.
- Rosenberg, B. (1973). Linear regression with randomly dispersed parameters. *Biometrika*, 60, 61-75.
- Saka, T. (1989). *Indicators of school effectiveness: Which are the most valid and what impacts upon them?* American Educational Research Association, San Francisco. (ERIC ED306277).
- Shavelson, R. J., McDonnell, L. M., Oakes, J., & Carey, W. (1987). *Indicator systems for monitoring mathematics and science education*. Santa Monica, CA: Rand Corporation.
- Southern Regional Education Board. (1990). Incentive programs for teachers and administrators: How are they doing? *Career Ladder Clearinghouse*. Atlanta, GA: Southern Regional Education Board.
- Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. L., Merriman, H. O., & Provus, M. (1971). *Educational evaluation and decision-making*. Itasca, IL: Peacock.
- Town, S. W. (1973). Action research and social policy. *Sociological Review*, 12(4), 128-137.
- U. S. Department of Education. (1991). *AMERICA 2000: An education strategy. Sourcebook*. Washington, DC: Author.
- Webster, W. J. (1991). An analysis of available student achievement data in the Dallas independent school district. *Executive Summaries of Evaluation Reports*, 1-44. Dallas: Dallas Independent School District.

- Webster, W. J., Mendro, R. L., & Almaguer, T. (1994). Effectiveness indices: A "value added" approach to measuring school effect. *Studies in Educational Evaluation*, 20, 113-145.
- Webster, W. J., Mendro, R. L., & Ortiz, M. (1994). *Identifying effective schools: An empirical comparison of selected regression and hierarchical linear models*. American Educational Research Association, San Francisco.
- Webster, W. J., Mendro, R. L., & Almaguer, T. (1993). *Effectiveness indices: The major component of an equitable accountability system*. American Educational Research Association, Atlanta, Georgia.
- Webster, W. J., & Olson, G. H. (1988). A quantitative procedure for the identification of effective schools. *Journal of Experimental Education*, 56, 213-219.
- Webster, W. J., & Schumacher, C. C. (1973). A unified strategy for systemwide research and evaluation. *Educational Technology*, 13(5), 68-72.

5

AN ANALYSIS OF ALTERNATE MODELS

Introduction

As the Preamble to Chapter 4 has stated, this chapter will attempt to summarize presented evaluation models and contrast them based on three different, but related, ways of looking at them:

1. A summary display of the purposes of each model—which is placed under one of three headings: formative; formative and summative; and summative
2. An examination of each model aligned against the principles contained in the Joint Committee's *Standards* to determine main areas of strengths or weaknesses
3. A discretionary value judgment about the worth of each of these models aligned against the main uses of teacher evaluation models for decision-making

It is hoped that, taken together, these elements will comprise a user's guide. Such a guide stems logically from the development of the first three of this book's four main cores: standards for teacher evaluation, the Guide to improving teacher evaluation systems by applying the Joint Committee's *Standards*, and the presentation of alternative models for teacher evaluation.

The purpose of this chapter is not to place a misconception in readers' minds that any *one* formative or any *one* summative model will be selected as the answer to teacher evaluation. We have stated in earlier writing (1985, p. 6), our belief that if an audience is composed of consumers who need to choose a product or service,

then alternatives should be presented, together with information on how they compare on critical criteria. Moreover, it is most unlikely that any one approach to teacher evaluation will suit all schools and school districts with their varying contextual complexities, including students' diverse educational needs. Also, practice has clearly indicated that individual districts are adopting an eclectic approach to teacher evaluation, trialing and often adapting models to formulate an approach that responds best to the district's goals and particular circumstances. Often, however, in the eclectic process, one or two models are emphasized, with parts of others incorporated as needed.

The outputs of different evaluation models may be used formatively or summatively or both. Some models specify only one use, or are keyed to one use, while others are constructed to use both formative and summative roles of evaluation. Thus, as the next section indicates, we have divided the models by orientation: formative only; formative and summative; and summative only. We also ask readers to note that it is problematic to have the same person conduct both formative and summative personnel evaluations. In fact, our stance is that if at all possible, different people should undertake these evaluations. Sometimes, as in the case of a private school, the principal will have to be the sole evaluator. However, in a school district, it should be possible, as McGreal has pointed out in Chapter 4, to separate administrative (summative) from supervisory (formative) behavior. While it is never permissible for principals and other administrators to escape the responsibility of ascertaining whether a teacher is performing according to a district's requirements, it is possible, with common sense guidelines, for school administrators to act more as instructional supervisors for a teacher's professional growth than as a regulator of a district's objectives. Teachers should be evaluated both formatively and summatively, and the same criteria of judgment will apply. However, the purposes of evaluation may differ and so, too, if possible, should the persons carrying out the evaluations.

A Summary of the Purposes of the Ten Models Selected in Chapter 4

Tables 5-1, 5-2, and 5-3 give a tight synopsis of the ten presented models. Scriven (1967) defined the two roles of evaluation: *formative*, in which information and judgments are reported for developmental purposes, and *summative*, in which a judgment is given based on accumulated evidence about the extent to which the stated needs of consumers are met. Table 5-1 displays those models which are mainly formative, Table 5-2 those that have both formative and summative elements, and Table 5-3, those that are summative.

Table 5-1. Purposes of Formative Evaluation Models for Teacher Evaluation

Role of Evaluation	Formative				
	Model Developer	Hunter	McGreal	Iwanicki	Withers
Main Intentions	To improve teaching performance based on observation of teaching behavior and advice offered by trained observers providing feedback on strengths and weaknesses.	To improve teacher instruction as part of strengthening the education system continuously, through cooperation between evaluators and teachers, with focus on delivery system of instruction.	To improve teacher performance through the process of contract plans evolving in a cooperative, 5-stage process between teacher and evaluator. To enhance school effectiveness through an integrative response of staff development, school improvement, and teacher evaluation.	To encourage ongoing self-evaluation as a professional responsibility benefiting teacher and learner, based on rigorous standards, and to include evaluation by colleagues of each others' performance.	
Other Intentions/Outcomes	To develop professionally both teachers and observers (in respect to leadership responsibility). To influence teachers from novice to expert status. To develop a district's or school's skills in videotaping of teaching/learning situations.	To help teachers attain established performance expectations and to improve student learning in a planned, constructive way. To develop criteria for teacher effectiveness, and (individual) teacher goal-setting.	To employ performance review, to identify priority areas for improvement, to develop performance objectives, and, after implementing the improvement plan, to evaluate the impact on teacher performance. To contribute to organizational as well as teacher needs.	To enable teachers to plan future teaching on reflection of what has worked well. To make self-assessment holistic, integrating the evolving performance of teachers and students. To strengthen case for internal, rather than external, teacher evaluation.	
Proponents	District instructional leaders (from ranks of skilled teacher, curriculum experts)	Supervisor (as evaluator)	Evaluator Teacher	Teacher Colleague	
Beneficiaries (Direct and Indirect)	Teachers Students Schools Districts School Communities	Teachers Students School	Teachers Students School	Teachers Students School Society	

Table 5-2. Purposes of Formative and Summative Models for Teacher Evaluation

Formative and Summative				
Role of Evaluation	Model Developer	Manatt	Toledo School District	Shinkfield
Main Intentions		To improve teacher performance and influence student achievement (K-12) through skilled observation and lesson analysis. To change teacher behaviors based on record keeping for future decisions.	To organize teacher evaluation based on teacher and their professional organization, determining standards for presentation to education authorities, and to have trained, skilled teachers becoming the evaluators of beginning and failing teachers, with emphasis placed on both formative evaluation for professionalism (teacher improvement) and also on summative evaluation for accountability (including dismissal).	To develop teachers professionally through an intensive process of formative evaluation by ongoing, cooperative efforts by teacher, principal, and peer in a three-year cycle involving all teaching staff. To counsel poor performers out of teaching.
Other Intentions/Outcomes		To organize workshops for administrators/senior educators on school/teacher effectiveness approach to teacher evaluation. To develop teacher job targets, efficient time usage, rating scales, and due process. To include summative evaluation for improvement judged against predetermined standards.	To strengthen and maintain teacher morale through participative decision making and maintaining professional standards to offer counseling for both probationary and intervention program teachers. To remove from the principal the onus of evaluating intern and intervention teachers. To dismiss teachers not performing to a satisfactory standard if professional support and counseling fails.	To develop staff competency and duties list (and judgment criteria) that will be used for both formative, and, if needed, summative evaluations. To increase staff morale, principal-teacher trust, and, as the baseline, improved student confidence in learning.
Proponents		Principals, district administration, and teachers (involved with criteria selection)	Teachers, their professional organization, and administrators	Principal
Beneficiaries (Direct and Indirect)		Districts (TPE placed within this context) Community (the model has accountability elements)	Teachers Schools Principals Union and management District community	Individual teachers All staff Principal Students School

339

Table 5-3. Purposes of Summative Evaluation Models for Teacher Evaluation

Role of Evaluation	Summative		
	National Board for Professional Teaching Standards	Sanders (Tennessee VAAS)	Webster (Dallas ISD)
Main Intentions	To establish high and rigorous standards for what teachers should know and be able to do and to certify teachers who meet these standards. To advance other educational reforms for the purpose of improving student learning nationally and to restore public confidence in schools.	To assess impact of educational systems, schools, and teachers on the learning gains students make yearly on norm-referenced achievement tests; and to report individual teacher effects to teacher, administrators, and school board members for accountability purposes (thus making teacher assessment a function of student achievement).	To use standard performance outcomes and other variables as a basis for school and teacher evaluation (using the school and not the teacher as the unit for measurement) and to operationalize accountability through criterion-referenced (analysis of absolute outcomes) and norm-referenced (school effectiveness indices) methods.
Other Intentions/Outcomes	To recognize and reward the work of accomplished teachers and to increase the status of the teaching profession. To expand the flow of first-rate people into teaching, and to retain accomplished professionals.	To account for team-teaching, apportion accountability to teachers involved. To base teacher assessments on 3-5 years of data. To employ statistical analyses that endeavor to overcome traditional problems associated with using student achievement data in educational assessments.	To involve the district educational and school committees in planning. To give financial rewards to schools and staffs based on accountability measures (with allowances made for differing school populations). To foster inter- and intra-school cooperation by focusing on important schooling outcomes and tracking progress of individual students.
Proponents	Board of Directors of NBPTS and the staff	Tennessee Department of Education Researchers and statisticians	Dallas Independent School District Researchers and statisticians
Beneficiaries (Direct and Indirect)	Teachers Students Schools and districts Nation	Students (the stated aim is to improve educational opportunities) Schools (if sufficient useful information is generated by TVAAS)	Teachers and schools that perform well on accountability indices Students (if outcomes of recording progress are used effectively)

These three tables outline the model developer's main intention, together with less significant intentions (or outcomes) followed by a listing of the proponents of the model, (i.e., those who organize and implement the evaluative activity), and finally a list of those who benefit, directly and indirectly, from the outcomes of the teacher evaluation process.

Formative Evaluation Models

The four models by Hunter, McGreal, Iwanicki, and Withers displayed in Table 5-1 have both overlapping purposes (and often outcomes) and differing emphases, as the following brief discussion will indicate.

The Hunter model, like the Manatt, Shinkfield, and Iwanicki approaches, is theory based. Evaluations are conducted by trained observers, who provide teachers with feedback for improving teaching skills and performance so they can comply with what is theorized to be sound teaching. The Hunter model is centered on clinical supervision, an art form she perfected based on the work of the originator of this approach, Keith Goldhammer, in 1969. Reference has been made to his work in Chapter 4 (McGreal).

The McGreal model is concerned with the place and importance of teacher evaluation within an educational system (school or district) to strengthen that system. He, too, sees the basic value of an educational system developing minimum performance expectations and the development of criteria for teacher effectiveness in required duties and competencies areas. All four models lay some claims to responsiveness to consumer needs; one common factor is to help implement improved student learning.

Although the Iwanicki model has a provision for summative evaluation, its overwhelming emphasis is on formative evaluation to meet the professional needs of teachers. For his approach to succeed, the basic assumption is that teachers are professional people who look to improve their performance and thus to enhance student learning. Contract plans become the vehicle for this to happen.

Teacher self-evaluation plays an important part in the Hunter, McGreal, and Iwanicki models (also in the Shinkfield model and indirectly in the Manatt model, which are contained in Table 5-2). In fact, self-evaluation is a vital step in the partial fulfillment of one of the aims of these approaches, namely, that *teachers are responsible for their own development as professional people*. The Withers' model, which is centered around teacher self-appraisal, exemplifies this point. Withers underlines it by emphasizing functional standards that must accompany any thorough use of self-evaluation. Moreover, he strongly contends that colleagues must evaluate each others' teacher performance to help make self-assessment holistic, integrating the evolving effectiveness of teacher performance and student learning.

Formative and Summative Evaluation Models

Table 5-2 displays three models that have both formative and summative evaluation characteristics. The first, by Manatt, progresses (as Chapter 4 has shown) from formative to summative, with emphasis remaining on teacher improvement. Shinkfield's model is predominately formative, but provision is made for some outcomes (e.g., dereliction of duty) to be treated summatively. The Toledo School District model, by comparison with those of Manatt and Shinkfield, places considerably greater weighting on summative evaluation.

The Manatt model addresses the growing concern for the need for improved teacher performance by basing teacher evaluation on an analysis of progress made toward the accomplishment of predetermined objectives or job targets. Of the six stages in his model, Manatt considers four to be formative and two to be summative evaluation, which is viewed more as a mechanism for strengthening performance than as an instrument to dismiss poor teachers.

In his approach, Shinkfield bases effective outcomes, (i.e., improved teacher skills and student learning) on an intensive, ongoing formative evaluation requiring strong trust between teacher and principal. However, the Staff Competency and Duties List is a basis for both formative and summative evaluations. And the professional responsibility of both teacher and principal in respect to this list (as the teacher must meet minimum standards) requires outcomes to be viewed both formatively and summatively, even though the model gives particular emphasis to formative evaluation.

The general orientation of the Toledo model is on the enhancement of teacher development. To this end, probationary teachers and those who are failing to meet minimum standards (and who enter the intervention program) are counseled by skilled, trained teachers. This approach is formative. Summative evaluation is used to dismiss teachers who cannot perform at the required level. While most unusual circumstances gave birth to this system of teacher evaluation and for this reason it may be difficult for other districts to implement, its underlying principles of determined participative decision making are well worth examining.

Summative Evaluation Models

The three summative teacher evaluation models displayed in Table 5-3—those of the National Board for Professional Teaching Standards, William Sanders' Tennessee Value-Added Assessment System, and William Webster's Value-Added and Product Measures approach for the Dallas Independent School District—are more distinctive from each other than the four formative models, or the three forma-

tive/summative approaches. One common element is accountability. But even here, the meaning differs somewhat among the models. One intention of the National Board is to make teachers more accountable to their profession by gaining a nationally recognized qualification to the advantage of education and to the students they teach. In Tennessee, the Department of Education is held legislatively responsible and accountable for improved student learning. And in the Dallas model, accountability at the administrative level becomes the responsibility of the District administrators and the District's schools. In all three models judgments are made on the basis of accumulated evidence about the degree to which stated needs are met. These needs may be nationwide (as is the case with the work of the National Board) statewide (Tennessee), or districtwide (Dallas).

The summative nature of the National Board's efforts is contained in its principal aim: *to establish high and rigorous standards for what teachers should know and be able to do, and to certify teachers who meet those standards*. The standards are being devised on the basis of what comprises accomplished teaching, with the longer-term aim to strengthen the nation's schools and their outcomes by improving teacher quality.

In the Tennessee model, the outcomes are used summatively for accountability purposes. As Chapter 4 has shown, the model assesses the impact of educational systems, schools, and teachers on the learning gains students make yearly on norm-referenced achievement tests. Results are reported for accountability purposes. It should be noted, however, that Tennessee still uses the career ladder in addition to the Value-Added System, thus offering a broader dimension overall to teacher evaluation.

It should be noted that in Dallas, although a record is maintained of each student's record of progress (in a number of developmental areas), the effectiveness of the teacher and other staff in a school is the unit of measurement. Dallas operationalizes accountability summatively through criterion-referenced (analysis of absolute outcomes) and norm-referenced (school effectiveness) methods.

Neither the Tennessee nor the Dallas schemes have yet developed to the stage where a critical analysis can fairly be made about either. Both, nonetheless, have placed such extensive resources into the development of their models that many educational and administrative leaders are awaiting the outcomes of these notable evaluative enterprises. Perhaps the important point to note is that summative assessments of major aspects of schools, aided by massive computerization, have now become a reality.

An Examination of the Models Against the Joint Committee's Standards

Each of the ten models was evaluated against the Joint Committee's *Standards* to find main strengths and weaknesses. As has been stated on other occasions in this book, it is unlikely that any one model will perfectly meet all the needs that arise from the particular context of a district or school; nor, indeed, are there any perfect models. Nonetheless, a listing of perceived strengths and weaknesses may assist readers to select among, or improve upon, existing models along the lines suggested in Chapter 3, School Professionals' GUIDE To Improving Teacher Performance Evaluation Systems.

Complete details of the analysis of each model against the *Standards* have not been recorded here, since we consider it sufficient to inform readers only of the more important conclusions reached about strengths and weaknesses.

Hunter Model

Main Strengths

- The model promotes sound educational principles.
- Guidelines are clearly articulated.
- Evaluatees are always addressed professionally and constantly encouraged.
- Evaluatees are assisted toward achieving the aim of providing excellent services.
- Reporting is timely, practical, and appropriate.
- All concerned parties are constructively involved.

Main Weaknesses

- The process is costly in terms of both time and money.
- The context of the evaluation, the classroom, is insufficiently recorded to identify constraints on performance.
- There is no provision for safeguards against bias (although stringent training of supervisors could obviate this problem).
- It is a limitation that the model is devoid of a stated concern for evaluation (although formative evaluation is implied).

McGreal Model

Main Strengths

- Effective performance of job responsibilities is stressed.
- All appropriate records are made of evaluation policies and practices to allow evaluations to be equitable and in accordance with relevant laws.
- Emphasis is given to a strong professional relationship between evaluatee and evaluator; self-esteem is of paramount importance in this model.
- Development of teachers is based on constructive, well-planned strategies.
- Follow-up helps understanding of outcomes and gives impetus for appropriate changes.
- All interested parties are closely involved; sound documentation is demanded at each stage of the process.

Main Weaknesses

- No assurance is given about reports being confidential to legitimate users.
- No emphasis is given to the necessity to provide adequate resources to implement the model.
- Insufficient provision is made to assure reliability and systematic data control.
- No place is given for periodic review of the model for revision and strengthening.

Iwanicki Model

Main Strengths

- Strong considerations are given to the welfare of evaluatees.
- Work plans emphasize the importance of effective teaching performance and adherence to job responsibilities.
- Guidelines are recorded as policy after negotiated agreements have been reached to allow evaluations that are consistent and equitable.
- The self-esteem and motivation of evaluatees are enhanced by the professional nature of the evaluation.
- The model guides and assists those evaluated to provide increasingly valuable service.
- Follow-up procedures allow evaluatees to have greater understanding of results and to take appropriate actions.

- Documentation during or at the end of the process indicates the extent to which work plan objectives have been realized.

Main Weaknesses

- The model places so much weight on teacher improvement that it lacks credibility with respect to identifying serious teaching deficiencies.
- Insufficient heed is given to limiting access of final reports to those most closely involved and concerned with outcomes.
- The model cannot safeguard against a number of organizational factors that may affect the required timeliness of reports.
- Despite emphasis on evaluator training, evaluator bias control cannot be assured, raising the possibility of the evaluatee's performance not being fairly and objectively assessed.

Withers Model

Main Strengths

- The model strives to fulfill both institutional and personal goals through enhanced teacher performance.
- Emphasis is given to teacher self-encouragement helped by collegial assessment of both teacher and student learning improvement.
- All planned activities lead toward the teacher taking appropriate actions.
- The model has the advantage of parsimony of time and resources, although adequate teacher and colleague time is essential for successful implementation.

Main Weaknesses

- No method is suggested to help ensure that teachers have the necessary skills and confidence to be involved in self-evaluation.
- Reporting too easily can be haphazard and undirected; methods to strengthen the practical value of reporting, so important to self-evaluation, also are needed.
- The main validity problems center around measurement procedures (What is it that is really being assessed and how?); similarly, reliability is not assured.

Manatt Model

Main Strengths

- The model aims at fulfilling both institutional missions and teacher development.
- Guidelines, including policy decisions and negotiated agreements (about the process of teacher evaluation) are mandatory.
- Evaluatees should perceive an enhanced attitude toward evaluation and its purposes.
- Emphasis is given to evaluator credibility.
- Follow-up is stressed in job target formation.
- Collaborative elements run through the process.
- The role of the evaluatee and the evaluator are clearly defined.
- There is direct emphasis on assessing the teacher's contribution to student learning.

Main Weaknesses

- Although the model is designed with a summative "top," its function (teacher improvement) is essentially formative.
- There is no apparent safeguard against bias to ensure teacher performance is fairly and objectively assessed.

Toledo Model

Main Strengths

- Guidelines have been scrupulously developed through collaborative efforts of main groups involved in the process.
- Through union, teacher, and administration discussions, conflicts of interest are solved as part of planning so that evaluation outcomes are not compromised.
- Both users and intended users of the model are clearly identified.
- The teachers who are evaluators are carefully selected and thoroughly trained so that evaluator acceptability and credibility are maintained at a high level.
- The model demands well-planned and well-executed procedures, culminating in timely and explicit summative reports for appropriate action.

- The feasibility standards are all very well met by this model: practical procedures, collaborative involvement of all concerned parties, and provision of adequate resources to implement the scheme.
- Methods are in place for the system review (and revisions have resulted).
- The model clearly defines the roles, responsibilities, and qualifications of the evaluator and the performance objectives of the evaluatee.

Main Weaknesses

- The model is expensive in terms of teacher time (and therefore finance).
- Insufficient emphasis is placed on the importance of student learning in the evaluation process.
- There is no apparent heed given to classroom influences and other constraints on teacher performance.
- With different teachers acting as evaluators (however well trained), there is the danger of reliability and validity being violated.

Shinkfield Model

Main Strengths

- The model emphasizes professional development through positive appraisal techniques based on promoting improved teaching and learning.
- Guidelines are contained in a collaboratively developed (staff and principal) handbook.
- Confidentiality is assured.
- Professional esteem and reputation of the teacher being evaluated are maintained.
- Evaluative procedures are aligned to a developed teacher competency and duties list.
- Verbal and written reports are timely and of immediate practical value and are supported by follow-up procedures.
- Documentation procedures are carried out thoroughly.
- The emphasis is on an in-depth study of the teacher over an extensive period of time.

Main Weaknesses

- The process is expensive in terms of principal (or administrator) time.

- No provision is made for conflicts of interest (which might affect the evaluation process).
- The classroom context is not defined, thus allowing environmental influences to confound perceived teacher performance.
- The model allows insufficient reviewing of its processes for improvement purposes.

NBPTS Model

Main Strengths

- The National Board has sought communication and collaboration with a wide range of persons, associations, and other organizations vital to the success of this national venture.
- The National Board's activities are aimed at promoting sound education principles, including raising the level of effective (accomplished) teacher performance and meeting student needs.
- Formal guidelines for accomplished teacher evaluations are thoroughly recorded in both policy and procedure forms ensuring, as far as possible, that evaluations are consistent, equitable, and in line with ethical codes of behavior.
- The development of the National Board's standards are designed to encourage and assist those evaluated to provide excellent service.
- Solid efforts are being made to plan and conduct assessment procedures so that they produce needed information while attempting to minimize costs.
- Strong resources have been provided to ensure that the National Board's activities are both effectively and efficiently implemented.
- The National Board has stated that its system will be reviewed periodically so that appropriate revisions may be made.

Main Weaknesses

- The validity of outcomes is highly suspect, as there is not a thorough, well-planned, and well-executed observation of candidates by credible evaluators.
- While assessments might reveal a rich array of teaching capabilities, scoring could prove a very real difficulty in such a major enterprise. Unless scoring is both valid and reliable, confidence in the entire scheme cannot be sustained.
- The very magnitude of the National Board's task makes it difficult for assurances to be given that assessments will be executed consistently by persons with the necessary qualifications, skills, and authority.

- Despite the stated attempts to keep costs to a minimum, certification requirements may prove too expensive for many teachers.

Tennessee Model

Main Strengths

- Testing of the model's assumptions showed that a minimum of three data points on each student produced teacher effect data that were not influenced by student characteristics.
- The model demands a very explicit description of the school and district culture to counter environmental influences and other constraints on teacher performance.
- Evaluation procedures, including statistical methodologies, are thoroughly documented so that users can assess actual outcomes against intended outcomes.
- Information used in the model is carefully processed and maintained to ensure systematic data control.
- The model aims to discover the impact of systems, schools, and teachers on the student learning gains and thus provides valuable information for administrative decision making, and also for state-level policy analysis.

Main Weaknesses

- The costs are such that only a very large district or state department could consider using the model.
- Decisions for the implementation of the scheme were by political fiat rather than by collaboration and negotiation between concerned parties.
- Access to teachers' evaluation reports seems to go beyond those with a legitimate need to review and use the reports. However, this is a function of state law rather than the evaluation model.

Dallas Model

Main Strengths

- This model (which is geared for use by large school districts) aims to fulfill sound educational goals and the institutional mission of accountability to meet the needs of students and the school community.
- As this model pits schools against each other in vying for a cut of merit pay money, it encourages teachers within a school to cooperate rather than compete.
- Guidelines, including policy statements, are collaboratively developed to give assurance of consistent, equitable, and legal procedures.
- If procedures are thoroughly planned in line with the intentions of the model, needed information will result cost effectively.
- The model indicates ways to define the district's and schools' roles and responsibilities in the evaluation system, including the place and importance of quality documentation.
- Measurement procedures are developed and implemented to assure validity and, in particular, reliability.
- The model is strong in all aspects of data control, and a systemwide account is given of the differences between schools on selected variables.
- The model explicitly considers a wide range of student outcome measures.
- A stakeholder accountability commission determines what outcomes are to be measured.

Main Weaknesses

- The model is expensive to implement and maintain and therefore could be considered only by a large school district.
- It is difficult to relate summative reports, for accountability purposes, to individual teachers and their comparative input into a school's tests and other results.
- Only two student data points are used, raising questions about reliability.

The Worth of the Presented Models for Decision Making

The final section in the Analysis of Alternative Models turns to an examination of the kinds of situations where teacher evaluation models should prove useful for decision making. There are many such situations, but as the focus of this book is

Table 5-4. Models' Response to Teacher Decision Situations

Decision Situation	Model									
	Hunter	McGreal	Iwanicki	Withers	Manatt	Toledo	Shinkfield	NBPTS	Tennessee	Dallas
Pretenure (Intern) Retention		✓				✓				
Professional Development	✓	✓	✓	✓	✓		✓	✓		
Tenure					✓	✓			✓	
Posttenure Retention					✓	✓	✓		✓	
Merit Pay or Similar Benefits					✓			✓	✓	✓
Promotion (or preparation for Area of Responsibility)	✓	✓	✓		✓		✓	✓	✓	✓
Reduction in Teaching Force					✓	✓			✓	✓
Dismissal (for Just Cause)					✓	✓			✓	✓
Self-Assessment and Self-Development	✓	✓	✓	✓	✓		✓	✓		

on the evaluation of *practicing* teachers, evaluative techniques to aid preteaching circumstances (e.g., selection and licensing) will not be addressed.

A major question raised in CREATE TEMP Memo One (September 1991) is whether it is appropriate to consider a *single* model of teacher evaluation as appropriate for all teacher evaluations (including preteaching, selection, and training). There is an argument favoring this, based on a working definition of the merit of a teacher. Any use of the same model for different circumstances, however, would imply different data sources, procedures, and possibly personnel being involved.

Traditional practice has tended, however, to consider using different teacher evaluation models for different types of decision making. Table 5-4 lists some of the critical decision-making stages during a teacher's career and indicates whether the selected models can respond usefully to that particular process. It should be stressed that a liberal interpretation of the possible applications of the model has been given in the sense that both direct and implied or potential uses of a model have been recorded (with a check mark) in Table 5-4. For instance, most of the models complying with the formative role of evaluation (see Table 5-1) directly or indirectly include self-evaluation; and some of the summative models, or variations of them, could be used for merit pay, promotion, reduction in teaching force, and dismissal decisions.

The contents of Table 5-4 need little embellishment. Between them, these models are able to offer help in all the decision-making situations listed. Once again, it is worth mentioning that a school or district may choose to develop a model or models eclectically, having considered their specific needs and contexts. Some of the models listed readily lend themselves to adaptation or inclusion as part of a model that offers wider opportunities for appropriate decision making. Moreover, as has been mentioned in the preamble to Chapter 4, Michael Scriven's duties-based approach to teacher evaluation could add strength to existing formative or summative models or both, and Hans Andrews' excellent advice about approaching dismissal (and similar) decisions could be included in the Manatt and Iwanicki models and others.

References

- Center for Research on Educational Accountability and Teacher Evaluation. (1991). *TEMP Memo 1*. Kalamazoo, MI: Western Michigan University.
- Scriven, M. S. (1967). The methodology of evaluation. In *Perspectives of curriculum evaluation*, (AERA Monograph Series on Teacher Evaluation, No. 1). Chicago: Rand McNally.

Stufflebeam, D. L., & Shinkfield, A. J. (1985). *Systematic evaluation*. Boston: Kluwer-Nijhoff Publishing.

INDEX

Accountability (see Educational accountability)
A nation at risk, 9, 23, 25, 44, 184
 Andrews, H., 174, 394

Bobbitt, C., 12
 Bolton, D. L., 13, 249
 Carnegie Task Force on Teaching as a Profession, 36, 44, 184
 Center for Research on Educational Accountability and Teacher Evaluation (CREATE), 1, 7, 8, 34, 83-84, 86
 National Evaluation Resources Services, 35
 Certification of teachers, 26
 Clinical supervision, 187
 Contract plans, 180, 245-249

Dallas Independent School District, 27, 383, 351-352, 355, 392
 An accountability system featuring value-added and product measures of schooling, 350-372
 Dwyer, C. A., 27, 46, 62-79

Educational accountability, 24, 182
 of a teacher evaluation system, 186, 350-374
 Educational Testing Services, 26-27, 66
 Evaluation (see also teacher evaluation)
 basic principles, 52-59
 development of technology, 1
 formative, 23, 313, 318, 378
 history of teacher evaln., 9-41
 legal and political aspects, 28-29
 summative, 23, 24, 318, 378
 teacher, 1, 43, 82, 106, 174, 209, 350-352
 Goldhammer, K., 213, 382
 GUIDE to Improving Teacher Evaluation, 5, 6, 81-153

Hunter, M., 177, 382, 385
 Instructional effectiveness through clinical supervision, 187-207
 lesson design, 188-189, 196-199
 observation of teachers, 199-202
 planning, implementing and evaluating the inservice program, 192-194
 preparation of leaders, 189-192

- TAI, (see Teacher Appraisal Instrument)
 teacher as decision-maker, 194-196
 types of supervisory conferences,
 202-206
- Inspectorial system U.K., 12
- Iwaniki, E.F., 175, 180, 382, 386
- A professional growth-oriented approach,
 245-260
 basic steps in evaln., 249-253
 contract plans, 245-249, 253-260
- McGreal, T. L., 175, 178-180, 382, 386
- Commonalities of successful teacher evaln.
 systems, 208-244
 characteristics of successful teacher
 evaluation (introduction), 208-209
 commonalities (eight) of successful
 teacher evaluation, 209-229
- McKenna, B., 82
- Manatt, R. P., 181, 382, 383, 388
- Developing a performance evaln. system,
 276-279
 judgment of teacher effectiveness, 287
 School Improvement Model Project
 (SIM), 181, 272, 274-276, 287
 teacher performance evaluation,
 271-290
 Teacher Performance Evaluation (TPE),
 181-2, 272-275, 278-287
- Models for Teacher Evaluation, 173-372
 definition of a model 173-175
 overview of alternative models 173-186
- National Board for Professional Teaching
 Standards (NBPTS), 27, 36, 184, 319-336,
 383, 390
 aims, 320
 assessment issues, 327-331
 benefits and costs, 333-334
 frameworks for certification, 324-326
 prerequisites for certification, 321-322
 standards and validation, 331-333
- National Commission on Excellence in Educa-
 tion: *A nation at risk* (see *A nation at risk*)
- National Education Association (NEA), 20
- Nevo, D., 82
- PRAXIS (see also The Praxis Series), 3, 26, 35
- Principal and peer evaluation of teachers
 302-319
 climate and policy, 305-307
 conferences, 307-308
 observations, 314-315
 responsibilities of the principal as evaluator,
 303-305
 Teacher Competency and Duties List, 309-
 311
 wind-up conference, 315-317
- Ryans, D. G., 19, 21
- Sanders, W. L. & Horn, S. P., 185, 383 (see also
 Tennessee Value-Added Assessment Sys-
 tem)
- School Improvement Model Project (SIM), 181,
 272, 274-276, 287
- School Professionals' GUIDE to Improving
 Teacher Evaluation Systems (also referred
 to as the GUIDE), 81-131
- Scriven, M., 4, 36, 86-88, 174, 394
- Shinkfield, A. J., 174, 183, 245, 302-319, 382,
 383, 389
- Standards: see The Joint Committee
- Standards (general), 32-33, 43-79, 331-333
- Professional teaching, 26
- Stemnock, S. K., 14-15
- Student Achievement Data in Educational As-
 sessment, 342-343
- Stufflebeam, D. L., 44, 45, 82
- Teacher Appraisal Instrument (TAI), 177, 187
- Teacher evaluation; (see also Evaluation)
 duties-based model, 35
 licensing, 50
 overview of teacher evaluation models,
 175-176
 preparation, 50
 principal and peer evaluation of teachers,
 183, 302-319
 professional development, 50

- rationale for evaluating teacher performance, 88
 - self assessment, 35
 - self-evaluation, 180, 261-271, 316-317
 - Standards*: see The Joint Committee
- Teacher Evaluation Models Project (TEMP), 34
- Teacher Evaluation Systems (see also the GUIDE), 98-100
 - alternative approaches, 120-121
 - documenting a system, 133-172
- Teacher Performance Evaluation (TPE), 181-182, 272-275, 278-287
- TEMP Memos, 4, 394
 - overview of teacher evaluation models, 175-176
- Tennessee Value-Added Assessment System (TVAAS), 27, 185, 337-349, 383, 391
 - general description, 340-342
 - philosophical underpinnings, 339-340
 - statistical mixed model, 344-348
- Student achievement data in educational assessment, 342-343
- The GUIDE, 81-131
 - key agencies, 102-104
 - organization for improvement, 128-130
- The Joint Committee. *The Personnel Evaluation Standards*, 3, 4, 6, 10, 31, 33, 44-46, 66-67, 81, 85, 88-94, 95-96, 104-105, 377, 385
 - applicability of the *Standards*, 59-61
 - applying the *Standards*, 105-128
 - Accuracy Standards*, 55
 - Feasibility Standards*, 55
 - Propriety Standards*, 52
 - Utility Standards*, 53
- The National Commission on Excellence in Education, 26
- The Praxis Series: Professional Assessments for Beginning Teachers (see also PRAXIS), 61, 63, 67-77
- Thomas, B., 82
- Toledo School District: Interim and Intervention Programs, 182, 289-301, 383, 388
 - interim and intervention programs, 297
 - responsibilities for evaln., 291
 - teachers as evaluators, 292
- U.S. Department of Education, Office of Educational Research and Improvement (OERI), 8, 24, 34
- Webster, W. J., & Mendro, R. L., 186, 383,
 - An accountability system featuring value-added and product measures of schooling, 350-372
 - District Improvement Plan (DIP), 356-362
 - school centered, 351
 - school effectiveness indices, 362-371
 - school improvement process, 353-356
- Withers, G., 180, 382, 387
 - Teacher self-assessment
 - autonomy and evaluation, 263-264
 - co-professionalism, 264-267
 - getting value from self-evaluation, 261-271
 - practical issues, 267-270
- Zelanak, M. J., & Snider, B. C., 16

TEACHER EVALUATION: Guide to Effective Practice is organized around four dominant, interrelated core issues: professional standards; a guide for applying the Joint Committee's *Standards*; ten alternative models for the evaluation of teacher performance; and an analysis of these selected models. The book draws heavily upon the research and development conducted by the federally funded national Center for Research on Educational Accountability and the Teacher Evaluation (CREATE). **TEACHER EVALUATION: Guide to Effective Practice** allows the reader to grasp the essence of the experience of sound teacher evaluation and apply its principles, facts, ideas, processes, and procedures. Finally, the book invites and assists school professionals and other readers to examine the latest developments in teacher evaluation.

ISBN 0-7923-9674-X



9 780792 396741

0-7923-9674-X